RESEARCH BRIEF

DEVELOPING A NEW MODEL OF TEACHER EFFECTIVENESS

Lessons Learned from the IEA Classroom Environment Study

Doris W. Ryan

This research project was funded under contract
by the Ministry of Education, Ontario.
It reflects the views of the author and not
necessarily those of the Ministry.

The Honourable Sean Conway, Minister
Bernard J. Shapiro, Deputy Minister

Ministry
of
Education
Ontario

87-00885/

DEVELOPING A NEW MODEL OF TEACHER EFFECTIVENESS

Lessons Learned from the IEA Classroom Environment Study

Doris W. Ryan

This research project was funded under contract by the Ministry of Education, Ontario. It reflects the views of the author and not necessarily those of the Ministry.

The Honourable Sean Conway, Minister
Bernard J. Shapiro, Deputy Minister

© QUEEN'S PRINTER FOR ONTARIO, 1986

Order Information:

Publications Sales
The Ontario Institute
for Studies in Education
252 Bloor Street West
Toronto, Ontario M5S 1V6

(416) 926-4707
Will invoice on orders over $30.00.
Other orders must be accompanied by a
cheque or money order payable to
O.I.S.E.

Publications Services
880 Bay Street, 5th Floor
Toronto, Ontario M7A 1N8

(416) 965-6015
(Toll Free) 1-800-268-7540
(Toll Free from area code 807)
Ask operator for Zenith 67200
Order must be accompanied by a cheque
or money order payable to the
Treasurer of Ontario

Contract 1051
ONO 2108

Canadian Cataloguing in Publication Data

Ryan, Doris W., 1934-
Developing a new model of teacher effectiveness

Bibliography: p.
ISBN 0-7729-1895-3

1. Teaching. I. Ontario. Ministry of Education.
II. Title.

LB1027.R92 1986 371.1'02 C86-090853-4

TABLE OF CONTENTS

Preface iv

I. Overview of Research on Teacher Effectiveness 1

- The Trait View of Teacher Effectiveness: Presage Studies and Presage-Product Studies 4
- Input-Output Studies 8
- The Behavior View of Teacher Effectiveness: Process-Product Studies 10
- Concluding Comments 14

II. Overview and Summary of the IEA Classroom Environment Study 15

- The Plan of the Study 17
- Participating Countries 18
- Variables and Constructs 19
- Instruments 21
- Data Preparation and Analyses 24
- Major Findings and Generalizations Derived from the CES 28
- Two Secondary Analyses of CES Data 37
- Concluding Comments 41

III. A New Model of Teacher Effectiveness Based on a Task View of Teaching 43

- Rethinking the Validity of the Process-Product Assumptions 44
- The Task View of Teaching 47
- A Framework of the Major Tasks of Teachers 48
- Implications of the Task View of Teaching 55
- Using Teacher Tasks to Link Research and Practice 58
- Summary and Concluding Comments 65

References 69

Appendix A 76

PREFACE

This monograph is written primarily for teachers, school administrators, educational policy makers, and teacher educators. Its central aim is to stimulate down-to-earth discussions, both of the assumptions we make about what makes one teacher more effective than another and of the actions and decisions we take as a result of these assumptions. Organizationally, the monograph: (1) presents a brief overview of research on teacher effectiveness, (2) provides an overview and summary of an international study of the relationships between teaching practices and classroom activities and student cognitive and affective learning, and (3) discusses the development of a new model of teacher effectiveness that grew out of our experiences in the international study. The first section provides the context for understanding the assumptions that provided the driving force for the IEA study, and the second section provides the context for understanding why my colleagues and I began to question these assumptions.

The monograph is not a technical research report of the IEA Classroom Environment Study. Thus, the text is devoid of the statistical analyses that are available elsewhere. The comprehensive report of the international study is currently in preparation under the editorship of Ryan, Anderson, and Shapiro. It will be published by Pergamon Press, likely in 1987. Interested readers may wish to refer to a special issue of the journal, Evaluation in Education (Ryan and Anderson, 1984), or obtain a microfiche copy of the report of Ontario's replication of the IEA study (Ryan et al., 1984), which is available from the Ontario Ministry of Education.

The summary of the IEA study that is presented here, and the generalizations derived from that study, reflect the hard work of several persons -- most notably, Lorin Anderson and Garrett Mandeville of the University of South Carolina and Norbert Sellin of the University of Hamburg. The model of teacher effectiveness discussed in the third section was developed cooperatively with Lorin Anderson and Robert Burns (of the Far West Laboratory for Educational Research and Development in San Francisco). I wish to express my deep gratitude to each of these colleagues.

D. W. Ryan

I. OVERVIEW OF RESEARCH ON TEACHER EFFECTIVENESS

Research on teacher effectiveness dates back to the turn of the century when the first recorded study on the topic appeared in print (Kratz, 1896; cited in Medley, 1982). Interest in such research escalated quickly and seems to continue unabated. More than 20 years ago, for example, Gage (1960) identified some 10,000 studies that had been conducted on the topic of teacher effectiveness. Even a cursory inspection of the more recent educational literature reveals that teacher effectiveness inquiry continues to attract the resources of researchers, funding agencies, and professional organizations (Doyle, 1978).

Why has interest in such research been so high for so long? Put simply, the societal stakes are very high. Education is one of the largest, if not the largest, of our social enterprises in terms of the number of people involved, the value of human time required in the schooling process, and the commitment of public funds (Walberg, 1986). The argument put forward two centuries ago by Adam Smith is still believable: education is a useful social investment because the wealth of nations depends upon the abilities of their people.

Parents and members of the general public are not likely to link their interests in schools and schooling to abstract theories of human capital. Moreover, they might not consider themselves to be supporters or adherents of research on teacher effectiveness or any other topic. However, they <u>are</u> concerned that children receive the best possible instruction in publicly-supported schools. There is evidence in articles published in the popular press that, despite what professional educators may consider to be improvements made in teacher certification requirements, teacher education, school curriculum and the like, many parents and lay persons continue to be critical of schools in general and teachers in particular. For example, a review of annual polls of the U.S. public's attitudes toward their public schools revealed that, for thirteen consecutive years, parents expressed reservations about teaching quality (Elam, 1978; cited in Bridges, 1986). Within Ontario, the public attitude polls conducted by OISE have also given voice to public concerns about the quality of education in our schools and classrooms (Livingstone et al., 1985).

Members of the education profession may not accept all of the criticisms put forward by parents and the general public, but they share the desire for quality instruction in our schools. For them, maximizing teacher effectiveness is a major goal (Medley, 1982). Although incompetent teachers likely represent a small proportion of the teaching force, they tarnish the image of the majority -- the competent teachers "who are overworked,

underpaid, and underappreciated for their efforts and accomplishments" (Bridges, 1986, p. 3). Neither school administrators nor teachers treat lightly the critics' demands for the removal of incompetent teachers from the schools. But what is the nature of teacher incompetence? How do school administrators distinguish incompetent teachers who should be dismissed from those who need assistance in expanding or improving their repertoire of competencies? The recent book by Bridges (1986), The Incompetent Teacher, demonstrates that these issues remain problematic. Professional educators are equally concerned with improving the performance of reasonably competent teachers. Thus, they seek answers to two additional questions: What characterizes effective teaching? How do school administrators, teacher educators, and professional organizations assist teachers in general to improve their effectiveness in helping students learn? Practicing educators justifiably expect assistance from researchers in answering all these questions.

What about researchers themselves? The interest they have exhibited in designing, seeking funding for, and conducting studies should not be attributed solely (or, perhaps, even largely) to desires to meet university standards for tenure. Certainly since 1970, a growing number of researchers view themselves as "improvers" and not merely "describers" of the teaching process, in Gage's (1966) terminology. Doyle (1978) cites several studies designed explicitly with the intention of providing sources of content for teacher education and "tools" that teachers may use to improve their instructional practices. These studies will be discussed shortly. Suffice to say that, within the educational research community, much of the research on teacher effectiveness would be classified as "applied research" or "improvement-oriented research" which is intended to produce knowledge useful to practitioners.

What knowledge has been produced about what makes some teachers more effective than others? And how has that knowledge been translated into practice? Before attempting to answer these questions, it will be useful to put forward some definitions of terms that will appear in the discussion. I have borrowed these definitions from Medley (1982), and I will quote him directly:

> The term "teacher effectiveness" will be used to refer to the results a teacher gets or to the amount of progress the pupils make toward some specified goal of education. One implication of this definition is that teacher effectiveness must be defined, and can only be assessed, in terms of behaviors of pupils, not behaviors of teachers. For this reason, and because the amount that pupils learn is strongly affected by factors not under the teacher's control, teacher effectiveness will

be regarded not as a stable characteristic of the teacher as an individual but as a product of the interaction between certain teacher characteristics and other factors that vary according to the situation in which the teacher works.

Evidence of the importance of these situational factors is found in the results of studies in which the achievement gains of two sets of classes taught by the same teacher (usually in successive years) have been compared. Measures of the effectiveness of the same teacher made in different classes have been found to correlate, on the average, only about .30 (Rosenshine, 1970; Veldman & Brophy, 1974). This indicates that the reliability of such a measure is only about .30, which implies that on the basis of student achievement alone, the teacher who is rated as most effective during one year is quite likely to be rated among the least effective during the next year.

"Pupil learning experiences" refers to the behaviors of pupils while learning is going on. This is not a teacher characteristic, but it has a great deal to do with how effective the teacher is, since the amount a pupil learns depends on what the pupil does (what experiences he or she has). Any effect the teacher has on pupil learning must result from some effect the teacher has on the pupil's learning experiences.

"Teacher performance" refers to the behavior of a teacher while teaching a class (both inside and outside the classroom). Teacher performance differs from teacher effectiveness. It is defined in terms of teacher behaviors, of what the teacher does, while teacher effectiveness is defined in terms of what the teacher's pupils do. Teacher performance resembles teacher effectiveness in that it, too, is a product of the interaction between certain teacher characteristics and the teaching situation. Teacher performance is often used as a basis from which teacher effectiveness can be inferred.

"Teacher competence" refers to the set of knowledges, abilities, and beliefs a teacher possesses and brings to the teaching situation. Teacher competence differs from teacher performance and teacher effectiveness in that it is a stable characteristic of the teacher that does not change appreciably when the teacher moves from one situation to another. It resembles teacher performance in that it has also been proposed as a basis from which teacher effectiveness can be inferred.

Teacher competence is conceived as a matter of repertoire. The knowledges, skills, and beliefs in a teacher's repertoire will be referred to as "competencies" that the teacher possesses. Depending on the purpose of the moment, competencies may be defined in terms of very small elements, such as "the ability to thread a sixteen-millimeter projector" or of very large ones, such as "the ability to implement the advance organizer model." A competency is defined in terms of what the teacher knows, believes, or can do. "Skill in classroom management" could be a competency. "Ability to manage a class" could not.

It should be noted that "competence" is a collective noun and has no plural, whereas "competency" does occur in the plural. It is

> customary to regard a competency as something that is either present or absent, but to regard competence as a matter of degree: the more competencies a teacher has, the more competent the teacher is.
>
> "Preexisting teacher characteristics" designates knowledges, abilities, and beliefs that a teacher is expected to possess on entering into professional training. For the most part, these are stable personality traits (like general intelligence or interest in children) that are believed to be relevant to successful teacher performance but that a teacher education program cannot and should not try to develop in students who do not already possess them.
>
> - Medley (1982, pp. 1894-5)

As we shall see, the research undertaken during various time periods has focused on different conceptions of teacher effectiveness. Moreover, the studies have had different impacts on educational policy and practice. Because practice tends to lag behind knowledge, recent research conceptions, assumptions, and findings have so far had limited influence. In contrast, the early research conceptions, assumptions, and findings still underlie educational practice and provide the basis on which most decisions related to teacher effectiveness are made. In Medley's (1982, p. 1894) words: "Totally invalid results have penetrated to the very core of beliefs of educational decision makers and affected them accordingly." We turn first to this early research.

The Trait View of Teacher Effectiveness: Presage Studies and Presage-Product Studies

For the first third of the century, a single research paradigm dominated inquiry into teacher effectiveness. The design was straightforward: elementary school pupils, or secondary school pupils, or teacher educators, or educators, or some combination of such groups were asked to try to remember the best teacher each of them had ever encountered and to write down what made that teacher different from others. The researcher then collated and compared these descriptions and derived, from them, a list of characteristics that supposedly distinguished effective teachers from ineffective ones. The list could then be used for judging whether teacher-candidates or practicing teachers would likely be or were "effective".

This kind of research design was repeated over and over again in a series of studies across more than 35 years. The studies differed from each other primarily in the variety of "judges" used and in the procedures used to produce the list or lists (Medley, 1982). The most extensive and sophisticated example of this body of research was the large-scale

Commonwealth Teacher Training Study (Charters and Waples, 1929), which produced a number of lists of varying lengths. Typical of the characteristics appearing on these lists were the following (the top six on a list of 25): adaptability, considerateness, enthusiasm, good judgment, honesty, and magnetism. At no point in this series of studies was any attempt made to check whether pupils taught by teachers perceived as possessing the characteristics did in fact learn more than pupils taught by teachers not perceived as possessing them (Medley, 1982).

Implicit in these early studies are two assumptions that, at least to some extent, are pervasive even today. The first assumption is that just about anyone who has ever been to school is a good judge of teacher effectiveness. Joyce and Weil (1972, p.3) observe that this assumption arises from the fact that, "As in the case of art, good teaching is something many people feel they can recognize on sight, although they have difficulty expressing a reasoned basis for their judgment." Medley (1982) notes that almost anyone forms a strong impression of how effective a teacher is after watching a teacher work, even if only for a few minutes; the phenomenon is universal. As noted above, the validity of such impressions, in terms of student learning, is not tested. Nevertheless, one must accept the assumption that good teaching can be judged subjectively if credibility is to be given to the approach to the study of teacher effectiveness used in the early research.

The second assumption is that "good" teachers are born, not made. This is implicit in the nature of the characteristics listed as a result of analyzing the impressions of the "judges" in these studies. Most of the characteristics would fall under Medley's definition of "pre-existing teacher characteristics." As such, they are not ones acquired by taking professional education courses. There is no evidence that either the judges (pupils or professional educators) who contributed to such lists or the researchers who compiled them make any distinction between these personality traits of teachers and any of the other types of variables related to teacher effectiveness. Thus, Medley concluded:

> This confusion among what are in reality quite different types of variables, which bear different relationships to how effective a teacher is or becomes, not only permeates the early research, but persists and muddies policy decisions based on these lists to this day. Many teacher educators and school administrators still base their decisions on these same subjective impressions (Medley, 1982, pp. 1895-6).

A few characteristics that Medley would consider to be "competencies" and "teacher performance" variables (either of which could, at least theoretically, be developed through teacher education programs) began to appear on "effective teacher" lists generated by

research in the mid-1930s, when interest in the "trait" research design was waning. While the first assumption mentioned above was still apparent, the second one was at least somewhat modified, although pre-existing teacher characteristics continued to dominate the lists produced by the studies.

One instructive study conducted at the end of this period was that by Hart (1936), in which high school pupils were first asked to list characteristics of the teacher each of them had liked best. The most frequently mentioned characteristics included on the subsequent list were teaching skill (a teacher competency) and five other characteristics (personality traits): cheerful; friendly; interested in pupils; impartial; and fair in grading and marking. These same pupils were then asked to think about whether the teacher from whom they learned most was the same as the one they liked best. Those who answered negatively were asked to describe how the most effective teacher differed from the most liked one. The list generated from this request was strikingly different from the first and included no pre-existing teacher characteristics. Rather, it consisted of one teacher performance variable (makes greater demands on students), two teacher competencies (has more teaching skills and has more knowledge of subject matter), and one aspect of teacher effectiveness (has better discipline), as that concept was defined earlier. Or course, the list of effectiveness characteristics was still based on subjective perceptions of teacher performance.

The use of subjective impressions to make decisions about a teacher's effectiveness is not the only legacy of the early studies. Another is the teacher rating scale, which appeared in one of the early yearbooks of the National Society for the Study of Education (Boyce, 1915). Its popularity grew rapidly. By 1930, one researcher (Barr, 1930) located 209 different rating scales that were being used to evaluate teachers or student teachers. The scales typically list a number of traits or characteristics similar to the lists described above and which represent someone's idea about what makes a teacher effective. The scales are used by classroom observers (usually supervisors) to record judgments of the degree to which the teacher is thought to possess each of the listed traits. The judgments typically are recorded on a numerical scale (from one to five or one to ten, for example). Often, the separate ratings are totalled or averaged to produce an overall indicator of teacher effectiveness. In this way, teacher effectiveness is inferred from ratings of teacher classroom performance.

Researchers were among the first to use this new device since it provided a convenient way to quantify observers' perceptions of classroom performance (Medley, 1982). To their credit, some researchers attempted to test the validity of the rating scales by calculating

the correlations between measures of the average achievement gains of pupils in a sample of classes and supervisors' ratings of the teachers of these pupils. The first such attempt was made in 1935 (Barr), and several attempts were made prior to around 1959. The results were the same: there was no correlation between supervisors' ratings of teacher effectiveness and measures of teacher effectiveness (in terms of student performance on an achievement test). Bridges (1986, p.8) reports the following representative conclusions drawn from such studies:

> ...superintendents, supervisors, and principals tended to rate good teachers low and poor teachers high (goodness defined by pupil growth in achievement)...Ratings by superintendents, supervisors, principals should not be accepted as the sole or valid criteria until persons in these positions have been re-educated for this responsibility (McCall and Krause, 1959).
>
> ...evaluations based on...supervisors' ratings and those based on measures of pupil growth and achievement were not significantly correlated (Anderson, 1954).
>
> Whatever pupil gain measures in relation to teaching ability it is not that emphasized in supervisory ratings (Jones, 1946).

Although no evidence has yet been published that supervisors' ratings provide valid indicators of teachers' effectiveness, the rating scale is still in widespread use in essentially the same form in which it was introduced (Medley, 1982). The use of such rating forms or scales has disturbing implications. If the supervisors' ratings have no validity, as the research clearly suggests, incorrect decisions are still being made about: (1) which students should graduate from teacher education programs; (2) which candidates will be certified as competent to teach (a practice recently introduced in some U.S. states); (3) which teachers will receive permanent certification; and (4) which teachers need inservice training or even ongoing supervision and evaluation for purposes of possible dismissal (Medley, 1982).

Perhaps equally disturbing is the fact that the use of a rating instrument by teacher educators and school administrators reflects what was, until fairly recently, the "state of the art" -- even though it was introduced in 1915! This is because, until fairly recently, researchers had not produced dependable knowledge of better ways to recognize effective teaching by observing it. Indeed, researchers also accepted observers' ratings as "proxies" for teacher effectiveness for a long period of time. Since their observers were used for research purposes, however, the observers received training relative to the definitions of criteria and training in making judgments relative to the criteria (characteristics listed on the rating scale of teacher effectiveness). There were, moreover, some cases in which the

ratings of teacher effectiveness that were made by trained observers were found to correlate with measures of student test performance (Rosenshine, 1971).

Growing out of the early "trait" research and the development of rating scales was a research paradigm that became known as the "presage-product" approach (Dunkin and Biddle, 1974). "Presage" variables typically included teacher characteristics and teacher training experiences (assumed to represent competencies); "product" variables typically were represented by some measure of pupil performance on a test, although the earliest studies of this type used observers' ratings as "proxies" of teacher effectiveness. I have already mentioned some studies that tried to link "presage" variables with student outcomes (products). Suffice to say here that these studies produced little reliable information about the relationship between teacher characteristics and student achievement (Dunkin and Biddle, 1974). A representative conclusion reached after a review of these studies by the Committee on Teacher Effectiveness of the American Educational Research Association was as follows:

> The simple fact of the matter is that, after forty years of research on teacher effectiveness during which a vast number of studies have been carried out, one can point to few outcomes that a superintendent of schools can safely employ in hiring a teacher or granting him (sic) tenure, that an agency can employ in certifying teachers, or that a teacher-education faculty can employ in planning or improving teacher-education programs (AERA, 1953, p. 657).

Input-Output Studies

Although input-output studies represent one tradition of research on school effects rather than on teacher effectiveness per se, I include them here for two reasons. First, their findings have sometimes been taken to support an assumption that teachers make little or no difference in student learning, an assumption diametrically opposed to the "teacher difference" hypothesis that underlies teacher effectiveness research. Second, the Classroom Environment Study that will be discussed in the second section of this monograph was conducted through the IEA and was the first IEA study that did not follow an input-output design. The IEA achievement surveys will be mentioned in the next section.

One of the most highly publicized of the input-output studies was the Equality of Educational Opportunity survey conducted by Coleman and his associates (1966). Barr and Dreeben (1978) observed that this survey marked a "watershed of research on school effects", stimulating a rapid expansion of inquiry based upon a production-function model

which was borrowed from the field of economics. In such research, the school is treated as a firm which transforms inputs (such as books, teacher characteristics, activities, physical plant and equipment) into outputs (such as increments of knowledge, change in attitude, or gains in achievement).

The Coleman study involved over 4,000 public schools and 645,000 pupils in the U.S. The study findings have been interpreted (or misinterpreted, according to Barr and Dreeben, 1978) as proof that schools have little effect upon student learning. Critics, however, pointed out that no classroom process variables were examined and no observations were made either in classrooms or in schools. As Averch and his colleagues (1972, p.4) commented, "Research in this approach views the school as a black box containing students... Resources are applied to the students in the box and from this application some output flows." In short, the production-function formulation ignores whatever the "firm" does to transform raw materials into outputs (Barr and Dreeben, 1978). A related criticism was that the research considered pupil achievement in relation to school differences rather than teacher (or classroom) differences. That is, the teacher, school, and student characteristics were all aggregated to the school level. Good, Biddle, and Brophy (1975) pointed out, however, that "If teachers make a difference, and if pupils are to be exposed to both good and poor teaching, then it should not be surprising that many schools are found to be quite similar in their effects on students."

Regardless of the criticisms of this approach and the paucity of research describing what takes place in classrooms, the input-output studies of the 1960s were taken to support a view that teachers are not differentially effective. In this view, students learn because of variables that have little or nothing to do with teachers or classroom environments -- variables such as student aptitude, interest in learning, and degree of parental support (see Stephens, 1967). Good, Grouws, and Ebmeier (1983, pp. 3-4) stated that:

> ...it was almost becoming a national sport to criticize teachers and schooling in the late 1960s and early 1970s. Popular books like Death at an Early Age (Kozol, 1967) or Crisis in the Classroom (Silberman, 1970), among many others, were frequently cited as evidence that schools or teachers made little difference in the educational advancement of students.

These authors also argue that an implicit argument in much of this literature was that there appeared to be little variability in school and classroom routines; "to see one classroom was to see them all". They argued that the proponents of the "no difference" conclusion should

provide evidence that (1) all teachers teach the same way or (2) that differential teaching behavior has the same influence on student outcomes.

It was in this context that the "process-product" research paradigm came to be used "as an organizing framework for most contemporary research on effective teaching" (Doyle, 1978, p. 165). The IEA Classroom Environment Study was based upon this research model, to which we now turn.

The Behavior View of Teacher Effectiveness: Process-Product Studies

An encyclopedia article by Mitzel (1960) marked the beginning of an important change in research on teacher effectiveness. Mitzel defined three classes of criteria of teacher effectiveness: presage criteria (teacher characteristics and teacher competencies), process criteria (teacher performance variables), and product criteria (the results teachers get or the progress of pupils toward some educational goal, i.e. the definition of teacher effectiveness given earlier in this section). He also added a fourth class of variables related to teacher effectiveness which he called "environmental" variables (later to be called "context" variables). They included a variety of situational factors that affect pupil learning but are not controlled by the teacher. Medley (1982, p. 1896) stated that Mitzel's article "appears to have been the beginning of a process of sorting out these variables so that researchers could study them separately."

Shortly thereafter, Gage (1963) wrote about an important development in measurement -- the introduction of "low-inference" observation schedules as an alternative to rating scales for measuring teacher performance. These schedules were called "low-inference" because they required the observer to make relatively little inference or extrapolation from what is objectively visible and audible in the classroom. Gage's paper also described a way of studying teacher effectiveness which came to be called "process-product" research. The design was as follows: (1) samples of the performance of each of a group of teachers were observed and the frequencies of predefined categories of behavior were recorded (occasionally these were supplemented with a rating scale to measure some aspect of teacher performance that was difficult to define in terms of specific categories of behavior); (2) a measure of the effectiveness of each of the teachers was obtained by adjusting the class mean (average) posttest score on a selected achievement test to account for differences in their scores on a pretest (by using analysis of covariance); and

(3) correlations were calculated between measures of teacher performance (the average frequency of occurrence of the designated behaviors, across all observed lessons, for each teacher) with the measure of average class achievement.

The decade of the 1970s produced, in the U.S., a large number of studies that followed the process-product paradigm. Crawford and Gage (1977) reviewed four of these studies, each of which was a relatively large-scale study of primary classrooms (Grades 1-3) and in which the "product" variable was a measure of pupil achievement in reading. The studies were by Brophy and Evertson (1974), McDonald and Elias (1976), Soar (1973), and Stallings and Kaskowitz (1974). The inferences ("teacher should" statements) that they derived were the following:

> Teachers should have a system of rules that allow pupils to attend to their personal and procedural needs without having to check with the teacher.
>
> Teachers should move around the room a lot, monitoring pupils' seatwork and communicating to their pupils an awareness of their behavior, while also attending to their academic needs.
>
> When pupils work independently, teachers should insure that the assignments are interesting and worthwhile yet still easy enough to be completed by each third grader working without teacher direction.
>
> Teachers should keep to a minimum such activities as giving directions and organizing the class for instruction. Teachers can do this by writing the daily schedule on the board, insuring that pupils know where to go, etc.
>
> In selecting pupils to respond the questions, teachers should call on a child by name before asking the question as a means of insuring that all pupils are given an equal number of opportunities to answer questions.
>
> With less academically oriented pupils, teachers should always aim at getting the child to give some kind of response to a question. Rephrasing, giving cues, or asking a new question can be useful techniques for bringing forth some answer from a previously silent pupil or one who says, "I don't know" or answers incorrectly.
>
> During reading-group instruction, teachers should give a maximal amount of brief feedback and provide fast-paced activities of the "drill" type.
>
> - Gage (1978, p.39)

It is interesting to compare these inferences with those derived by a review conducted by Medley (1977) of fourteen projects, most of which were large-scale, federally-funded studies of primary classrooms in which most of the pupils were of low socioeconomic status. The

"product" measures were scores on either mathematics or reading tests. Medley examined only those correlations that were found to be replicated in two or more different studies. These "dependable" correlations pertained to three aspects of teachers' classroom performance. The first aspect was the learning environment which the teacher creates and maintains. The more effective teachers' classrooms were ones in which the pupils were more orderly, the teacher was less permissive and spent less time "managing" the class, and the pupils received more praise and fewer rebukes. The second aspect was the use that teachers make of pupil time. The more effective teachers' pupils spent more time in academic activities and more time organized in a single large group with the teacher in charge, spent less time in small groups or working as individuals on seatwork, and were supervised more closely. These results have led to the definition of a variable called "academic engaged time" (Rosenshine and Berliner, 1978) or "academic learning time" (Fisher et al., 1980), which is closely related to pupil learning. Finally, the third aspect was the strategy that teachers follow in class discussions. The more effective teachers asked more questions and the pupils asked fewer questions; the teachers asked fewer "high cognitive-level" questions and more "low cognitive-level" ones; and the teacher was less likely to give feedback or to amplify or discuss pupil answers. Thus, the strategy preferred by the more effective teachers resembled drill or recitation much more than it resembled what is normally thought of as a discussion.

The results of yet a third review, this time of an even larger number of studies, were reported by Rosenshine (1976). He organized the findings into a strategy or model of teaching which he called "direct instruction." The concept "direct instruction" refers to academically focused, teacher-directed classrooms using sequenced and structured materials. Using direct-instruction techniques, the teacher's goal is to move the students through a sequenced set of materials or tasks by ensuring that: goals are clear to the students, time allocated for instruction is sufficient and continuous, coverage of content is extensive, performance of students is monitored, questions are at a low cognitive-level so that students can produce many correct responses, and feedback to students is immediate and academically oriented (Rosenshine, 1979).

Most importantly, these three major reviews all agreed that consistent, useful findings had emerged from the process-product research of the 1970s, and they also agreed rather closely as to the nature of these findings. Albeit limited to primary grades and, most often, to classrooms of so-called "disadvantaged" children, the findings in large part run counter to what most teacher education programs have presented to their students as the picture of an effective teacher. In Medley's (1982, p. 1898) words: "If undergraduates in teacher

education are still being told to be permissive, to have pupils spend large amounts of time in small, autonomous work groups, and to avoid large-group instruction, as they were in the past, it is time for teacher educators to take a hard look at this research."

Although the consistency of findings reported by these correlational studies seemed convincing, cause and effect relationships needed to be tested in experimental studies in which teachers were trained to use the behaviors indicative of "good" teachers and in which the effects on pupil learning were measured. By 1979, when the IEA Classroom Environment Study was designed, four such experimental studies had been published. These were by Anderson, Evertson, and Brophy (1979), Crawford et al. (1978), Good and Grouws (1979), and Stallings, Needels, and Stayrook (1979). These studies generally confirmed the results from the correlational studies from which their teacher training "treatments" had been derived. More importantly, they demonstrated that when teachers are trained to behave in these ways, the achievement gains of their pupils increase significantly (Gage and Fordham, 1980).

The process-product studies of the 1970s made important contributions to knowledge about effective teaching. They produced the "first clear-cut usable findings obtained in eighty years of research in teacher effectiveness" (Medley, 1982, p. 1898). Indeed, Medley concluded in 1979 that the findings and implications of these studies about task-oriented classroom climates, academic learning time, and direct-instruction processes should be made "imperative" for teacher education programs and practicing teachers (Medley, 1979).

Nevertheless, almost before the first results of these studies were published, the process-product paradigm was subject to some fairly strong criticism. The criticisms took several forms (see Gage, 1978, for a review and a rebuttal). Many of these related to the methods used in observing teacher behaviors and in handling these data (e.g. aggregating them across all observed lessons). I will discuss these criticisms in the third section of this monograph because the experiences we had with the IEA study provide illustrations of many of the critics' points.

It should be noted here, however, that the use of classroom observation and behavioral checklists by school administrators in teacher supervision and evaluation became quite commonplace in the late 1970s. This trend can be traced to the process-product studies and their use of low-inference observation systems. These tools of researchers are not suitable for use in teacher supervision, a point that I have tried to make when called on to give talks at professional development workshops (Ryan, 1981).

Concluding Comments

The literature review presented above was limited to the two major "schools of thought" which guided research on teacher effectiveness from the turn of the century to the late 1970s. I have called these the "trait view of teaching" and the "behavior view of teaching". Both have had an impact on educational practice, but the influence of the "trait view" likely has been more pervasive despite serious questions about its validity as an indicator of what makes one teacher more effective than another in enhancing student learning. The validity of the "behavior view" has also been questioned, as we shall see. The process-product researchers who based their research on the "behavior view" did, however, produce broad implications (or translations of their findings) that seem both valid as predictors of student achievement and useful in improving teaching practice.

The spirit of optimism that prevailed among researchers by the late 1970s was not dampened significantly by growing criticisms of the underlying view of teaching that was implicit in the process-product model. It was in this climate of optimism that the IEA Classroom Environment Study was designed and conducted. We turn next to a discussion of that study.

In the event, the IEA study commenced just at the time when the designs used in the early process-product studies were being improved. Moreover, new schools of thought were emerging about how best to conduct research in classrooms. The results of these improved process-product studies and of studies conducted from different starting points were not readily available at the time when the IEA study was designed. In effect, then, they form part of the "lessons" we learned as we struggled to understand our own findings.

II. OVERVIEW AND SUMMARY OF THE IEA CLASSROOM ENVIRONMENT STUDY

The International Association for the Evaluation of Educational Achievement (IEA), a voluntary consortium of research institutes from around the world, was established during the era of large-scale "input-output" studies in the 1960s. As we have seen, the results of many of these studies suggested that student achievement is influenced primarily by factors and variables outside of schools and classrooms. The results of a pilot study conducted in 1961-62 just prior to IEA's first comparative study of mathematics achievement, for example, indicated that "native ability" accounted for more of the explained variance in successful completion of a test item than did the teacher's "emphasis in class" (Purves, in press). These same results, with family background substituted for native ability, and with physical facilities, formal curriculums and teacher characteristics substituted for the teacher's "emphasis in class," were included in the report on Equality of Educational Opportunity (Coleman et al., 1966).

The initial studies conducted by IEA were comparative surveys of educational achievement in various subject areas. By 1970, the first results of these studies became available, and they suggested that only 40 to 50 percent of the achievement differences among students could be explained by the "input" variables included in the international designs (c.f. Thorndike, 1973; Comber and Keeves, 1973; Purves, 1973). Concerns were raised in IEA, even before the results were published, about two deficiencies in the study designs: (1) they were cross-sectional rather than longitudinal studies and (2) they included no measures of actual teaching. Recall that little was known at the time about what takes place in classrooms since there were few studies that included classroom observations. In input-output studies, the schools and the classrooms within them were conceived of as "black boxes". As a consequence, Anderson (in press) concluded, "Attempts to characterize them resulted in the equivalents of statements that they were 'boxes' and that they were in fact 'black'."

Many of the IEA researchers shared the view that, if longitudinal studies could be designed in such a way that achievement was measured at two or more points in time, and if teacher

behaviors were observed during the time period(s) between those achievement measures, the additional variation in student achievement might be explained. Moreover, such study designs could yield much more knowledge about school-related factors that contribute to differences in student learning. Within the IEA community, however, researchers from various countries disagreed about the feasibility or usefulness of such demanding studies. Thus, although a meeting of interested researchers was held in Stockholm in 1972 and a proposal was prepared for a possible international study, the proposal lay dormant for several years. During those years, as we have seen, increasing numbers of studies of classroom teaching were conducted in the U.S., and the combined results of the studies published in the late 1970s suggested that there were "principles of teaching" (Anderson, Evertson, and Brophy, 1979), "teaching practices" (Crawford et al., 1978), and "behavioral profiles" (Good and Grouws, 1979) that were associated with, and likely influenced, student achievement. Thus, it seemed both possible and fruitful to attempt such a study at the international level.

A 1978 meeting in Israel brought together a small group of IEA researchers to discuss the various research models or approaches that could be used. The draft proposal that resulted from this meeting (Levin and Lewy, 1978) suggested an experimental study. They proposed that one group of teachers in each country be given training in classroom process variables that appeared, from earlier studies, to be related to student learning. A control group of teachers in each country would not receive the special training. The two groups of teachers and their classrooms would then be compared in terms of observed classroom processes and the average achievement scores of their pupils. The process variables which would be used in the training program consisted of the following five clusters: learning time, classroom questions, learning and instructional cues, feedback and correctives, and learning climate. Since these clusters had been identified from reviews of other studies, the proposal suggested that a correlational study was not needed before proceeding to an experiment.

Two meetings were subsequently convened by IEA to discuss the Levin-Lewy proposal. The first meeting was held in Berlin in 1979; the second was in Hamburg in 1980. The intended scope and nature of the proposed international study changed significantly as a result of discussions at these two meetings. There was a great deal of reluctance to move into an experimental study based on teaching behaviors and process variables identified primarily from U.S. studies. It was noted that there was no evidence to suggest that classroom processes in different countries were similar (or dissimilar, for that matter). There seemed to be too many unknowns to justify an experimental study until an initial descriptive study had been attempted and completed successfully. Thus, the group's consensus was in favor of

following a two-stage process-product design similar to the U.S. model, namely, a "naturalistic" correlational study followed by an experimental study.

In summary, the IEA Classroom Environment Study (CES) was conceived during a period of renewed optimism concerning the potential impact of schools and classrooms on student achievement. This spirit was captured by one of the original study participants from Australia:

> The identification of teaching behaviors that most influence student achievement and attitude development has been of central concern to educational researchers during recent years. The initial pessimism which surrounded much of this research endeavor throughout the early 1970s has been replaced by a feeling amongst researchers that now there exists the beginnings of a sound body of knowledge about the influence of teaching practice upon student learning (Fordham, 1983).

Professor N.L. Gage of Stanford was initially selected to be the Chairman of the International Project Council for the study, and Tamar Levin of Tel Aviv University was selected as the International Project Coordinator. For a variety of reasons, both Gage and Levin resigned these posts prior to the beginning of the implementation of the study. They were replaced by Bernard Shapiro as Chairman and Doris Ryan as International Project Coordinator, both of OISE.

The Plan of the Study

The IEA Classroom Study had two primary aims. First, it was intended to identify the similarities and differences in teaching practices and the conditions of learning as they occurred in a variety of countries. Second, the results of the descriptive study were to be used to identify those teaching behaviors that were associated with greater student achievement. To the extent that such behaviors could be identified, they were to be combined to form more general teaching practices which could be taught to teachers in an experimental study, in an effort to improve both teaching and, therefore, learning. In the event, only the correlational study was conducted, and the discussions which follow are limited to the descriptive, phase one study.

The basic design of the correlational study was to be longitudinal. Each participating national center was to select one or more learning units that required a specified period of

time to teach to the students. The time period between the administration of a pretest at the beginning of the unit(s) and the administration of a posttest at the end of the unit(s) was to be a function of the number and type of units selected by that particular national center. In the Ontario replication in English language schools, for example, the pretest was given at the beginning of the school year and the posttest was given at the end of the school year; thus, the entire Grade 8 curriculum in mathematics was to be covered. In contrast, the Republic of Korea selected a single unit (or curricular topic) in Grade 5 science for its study; thus, only two months' time separated its pretest and posttest. During the time that elapsed between the pretest and posttest in each country, each classroom in the sample was to be observed from six to ten times at regular intervals. Data pertaining to instructional and management behaviors used by teachers and data about the participation of students in the instructional activities were to be gathered during these observations. Data on community, school, classroom, parent, teacher and student characteristics were to be collected at various times before and after the classroom observations. Finally, analyses were to be carried out to determine the nature and extent of the relationships between observed teaching practices and students' achievement and attitudes within the context of the community, school, classroom, parent, teacher and student characteristics.

Although all participating countries followed the basic design outlined above, they were allowed to make a number of important choices. First, each national center was allowed to choose the subject matter and the age or grade level of the students to be included in their study replication. Because of this choice, and unlike all other IEA studies to date, each national center also was allowed to design or select the pretest and posttest to be used to assess student achievement in their study. As noted above, the countries also could select the amount of curriculum that would be taught during the time between the pretest and the posttest. Despite these national-level choices, however, each replication of the IEA study was to collect information on a standard set of variables by using at least twelve instruments that were developed at the international level.

Participating Countries

Nine countries participated in the study, with three replications being conducted in Canada. One study in Ontario included classrooms in English-language schools, a second study in Ontario included classrooms in French-language schools, and the third Canadian study was conducted in French-language schools in the province of Quebec. For purposes of the international analyses, however, each of these studies was considered as a separate

"country". In addition to Canada, the other participating countries were Australia, the Netherlands, the Federal Republic of Germany, Hungary, Israel, the Republic of Korea, Nigeria, and Thailand. These countries, located on five continents, provided a culturally diverse data base for the international study.

Table 1 presents information about the studies conducted in the participating countries with the exception of the Federal Republic of Germany. The studies in all countries except the FRG were conducted during the period 1980-1983; Germany's study was conducted during 1984, too late to be included in the international data analyses.

As Table 1 shows, two countries selected Grade 5 mathematics for study, and five selected Grade 8 mathematics. The remaining countries chose fifth grade physics, fifth grade general science, and sixth grade ancient history. Although the study guidelines had recommended a minimum of 40 classes of students per country, it was not possible for all the countries to meet this target sample size. The number of classes included in the various replications ranged from 18 to 87; thus, the number of students per country also ranged widely, from just over 400 to more than 2500. The length of the study (that is, the time between the student pretest and posttest) ranged from two months to nine months. Finally, the majority of countries conducted eight observations per classroom. Some classrooms in one of the countries were observed only once, whereas in another country, ten observations per classroom were conducted.

Variables and Constructs

The first major challenge in planning the CES was to develop the list of teaching behaviors that would be included in the descriptive study. The participating researchers agreed that, to be included, a teaching behavior would have to have a demonstrated relationship with student learning outcomes (as evidenced by the findings of previously-conducted research), be amenable to definition and measurement in an international study, and be able to be taught to teachers in a relatively short period of time. (The latter criterion was selected because of the intentions to proceed with an experimental study following the completion of the correlational study.) From a variety of potential behaviors identified in several literature reviews which had been conducted during the 1970s (Crawford and Gage, 1977; Dunkin and Biddle, 1974; and Rosenshine, 1976, 1979), the cooperating IEA researchers selected behaviors that appeared to fit the criteria. These behaviors were then categorized as either management practices (intended to bring students into contact

TABLE 1

OVERVIEW OF THE EXECUTION OF THE CLASSROOM ENVIRONMENT STUDY

Country	Grade Level	Subject Matter	No. of Schools	No. of Classes	No. of Students	Study Length	No. of Observations
Australia	5	Maths	39	75	1963	3 mos.	8-10
Canada (Ont/Eng)	8	Maths	18	27	757	7 mos.	8
Canada (Ont/Fr)	8	Maths	14	18	421	6 mos.	8
Canada (Quebec)	7-8	Maths	N/A	30	749	6 mos.	10
Hungary	5	Physics	N/A	39	1180	N/A	5
Israel	6	History	11	22	671	4 mos.	1-5
Korea	5	Science	15	45	2400	2 mos.	15
Netherlands	8	Maths	17	50	1125	7 mos.	8
Nigeria	8	Maths	35	35	1416	N/A	4
Thailand	5	Maths	77	87	2572	9 mos.	6

with relevant learning tasks) or instructional practices (intended to directly influence the learning processes of students during their contact with the learning tasks). As well, sets of community, school, classroom, student, parent, and teacher variables were selected in order to provide a more complete understanding of the context within which teaching was to be observed. These decisions resulted in the inclusion of some 160 variables, about which information would be collected from several different types of instruments. (In fact, since each response to a question or to an item in a scale could potentially be a separate variable, there were at least 240 discrete variables initially.) Appendix A provides a list of the variables actually used in one or more analyses.

Because of the large number of variables to be included in the study, it was necessary to develop some type of framework to organize the variables in a logical manner, especially for purposes of the multivariate analyses to be performed. The study's steering committee chose to place the variables within fifteen groups or categories called constructs. Five of the constructs included what traditionally are referred to as demographic variables. These constructs were labeled "community characteristics", "school characteristics", "teacher characteristics", "home characteristics", and "student characteristics". Another five constructs included the selected classroom behavior variables, organized around the function the behaviors served (or perhaps more appropriately were expected to serve) in the classroom. For example, the function of some teacher behaviors or classroom activities is to teach knowledge or skills to the students. These behaviors and activities were placed in a construct labeled "teaching students." Similar clusters of behaviors and activities produced constructs labeled as "orienting students", "managing students", "assessing and evaluating students", and "student participation" (where the function from the students' perspective is to become actively engaged in the classroom activities and the process of learning). Two constructs included variables associated with ways in which teachers and students viewed one another and the classroom; these were labeled "teacher perceptions" and "student perceptions". Two other constructs included variables concerning the nature of the classroom itself and the type and amount of instruction the students actually received there; they were labeled "classroom context" and "quantity of instruction". Finally, a construct labeled "student outcomes" included the variables of student achievement and attitude toward the subject being taught.

Instruments

A second major challenge facing the participants in the study was the development of instruments which could be used to collect data on all the variables mentioned above. In all, fourteen instruments were developed for use in the CES; of these, only two (student pretest

and posttest) were developed or selected by each individual country. The others were developed at the international level. The use of all but two of the international instruments was required of all participating countries. The exceptions were a Classroom Rating Scale and an Opportunity to Learn form.

Teachers participating in the study were asked to complete, or provide information for, five instruments. The Initial Teacher Questionnaire consisted of 24 questions about the teacher's background and responsibilities, the school, the class that was being observed, and about his or her autonomy in making a number of decisions. For the Teacher Interview, teachers were asked to read descriptions of six different management approaches and five different instructional approaches and select the two descriptions that best fit their own instructional and management approaches. During the Pre-Observation Interview, which was conducted prior to each classroom observation, teachers were asked to read several statements and indicate which statement best describes their major intention or instructional objective for the to-be-observed lesson and the main emphasis for the to-be-observed lesson. The Final Teacher Questionnaire consisted of 16 multi-part questions related to issues such as frequency of use of various instructional materials, use of various tests and projects for assigning grades or diagnosing learning deficiencies, and use of various teaching methods and styles. The Opportunity to Learn form, which was optional, was intended to elicit teacher judgments of the extent to which their students received instruction related to the knowledge and skills assessed by the various items on the student posttest.

Students participating in the study were asked to complete four instruments. The Initial Student Questionnaire consisted of ten questions related to student and parental background and 20 attitude statements with which students indicated the extent of their agreement or disagreement. The Pretest, developed nationally, was intended to assess the extent of a student's knowledge and skills relevant to the subject matter prior to the study. The Final Student Questionnaire consisted of 36 statements, including some related to the task climate of the classroom and some which were the same attitude statements included on the Initial Student Questionnaire. Again, students indicated their agreement or disagreement with the statement (26 items) or indicated that they believed the statement to be true or false (10 items). The Posttest, again developed nationally, was intended to assess the amount of learning that had been acquired by students during the time period of the study.

Finally, classroom observers were required to complete four instruments for each classroom observation. The Classroom Identification and Information Form was used to record the class enrolment, the number of students present, and the time when instruction began and

ended. The Snapshot instrument, used five times during each observed lesson, required the observer to record the type of classroom activity occurring (e.g. lecture, review, discussion, oral practice or drill, testing, reading at seats, written seatwork, laboratory work, non-academic or management, or other), the role of the teacher during this activity (e.g. interacting, monitoring, or uninvolved), and the number of students who appeared to be engaged in the activity. The Five-Minute Interaction Instrument (FMI), also used five times during each observed lesson, required observers to code and record the interactions that occurred during five-minute segments. Codes were to be entered for each of three dimensions: the context of the interaction (e.g. large group, small group, private interaction, monitoring, transition, or teacher uninvolved); the direction of the interaction (teacher to individual student, teacher to group, individual student to teacher, group to teacher, and teacher to other adult); and the nature of the interaction itself. For the latter dimension, observers first chose one of the following columns: instruction, questioning, responding, feedback, or management. Within the column, they would then choose the appropriate code. There were 28 categories or types of interactions across these columns (e.g. uses an example, asks an opinion question, gives a short response, says answer is wrong). The fourth observation instrument, the Individual Activity Record (IAR), was also completed five times during each observed lesson. On each of these occasions, the observers would check each of eight preselected students to ascertain how many of them were engaged in the classroom activity going on at that time.

In summary, the classroom observers in each country were required to use a standard sequence of three instruments -- the Snapshot, then the FMI, then the IAR -- five times during each observed lesson. For a 40-minute period, for example, there would be five sequences of eight minutes each. Five of these minutes would be spent coding on the FMI; the Snapshot and IAR were fit into the remaining three minutes. Obviously, the use of this observation system required intensive training. The observation system had been developed largely by researchers at Stanford University who had used similar instruments in other process-product studies conducted in the U.S. One of these researchers, Margaret Needels, trained the observation coordinators from the participating countries at workshops held in the Netherlands and in Belgium. The country coordinators then held similar training workshops in their own institutions for the persons who had been selected as potential classroom observers. Efforts were made to recruit observers who had some teaching experience. The training workshops lasted for seven days and included practice sessions in classrooms as well as practice coding while watching videotapes of classrooms. Only those observer-candidates whose codings were reliable and "valid" (as assessed by comparisons

with a criterion coding key) were selected to conduct observations for the study. Inter-rater reliabilities were also checked from time to time during the study.

There was another optional instrument developed for possible use by the classroom observers. This was the Classrom Rating Scale, which required observers to make overall ratings, after leaving the classroom, of the adequacy of a set of classroom variables (e.g. clarity of teacher presentation). Ratings from this form were not used in the international analyses.

Table 2 provides a summary of the instruments included in the IEA CES, the time when they were administered or completed, and the persons who completed them. Note that Appendix A contains a list of the study variables that were used in the international analyses, organized by construct. The instrument source of each variable is also given.

Data Preparation and Analyses

Although this discussion may not be of general interest, it will provide some information about the monumental tasks involved in organizing, cleaning, and editing the data, creating data sets for each country, and conducting several different kinds of analyses. The discussion, though brief and as non-technical as possible, illustrates the complexity of dealing with data based on minute-by-minute interactions in classrooms and of handling the variety of data generated by a process-product study.

I have already mentioned the large number of potential variables about which data were collected. Many of these were eliminated from consideration on a number of grounds (technical and conceptual); others were combined to form a scale or composite variable. The organizing constructs were helpful in reducing the number of variables as well. Readers interested in the criteria and procedures used to select variables for placement within particular constructs should read Anderson (in press). We were still left with vast amounts of data gathered in each of the participating countries. Consider, for example, that for the interaction variables, approximately 60 codes were made on each of three dimensions (context, direction, and nature of interaction) during each five-minute segment of classroom time. This was repeated five times per lesson. Where eight lessons were observed, there might be as many as 60 codes X 3 dimensions X 5 segments X 8 lessons = 7200 codes for one classroom in one country! Thus, some aggregations were done by the

TABLE 2

SUMMARY OF INSTRUMENTS INCLUDED IN THE IEA CES

Time Completed	Instrument and Abbreviation	Persons Completing
Prior to the Study	Initial Teacher Questionnaire (ITQ)	Teachers
	Teacher Interview (TI)	Teachers
	Initial Student Questionnaire (ISQ)	Students
	Pretest (PRETEST)	Students
Prior to Each Observation	Pre-Observation Interview (POI)	Teachers
During Each Observation	Classroom Identification and Information Form (CII)	Observers
	Snapshot (SNAP)	Observers
	Five-Minute Interaction Form (FMI)	Observers
	Individual Activity Record (IAR)	Observers
	Classroom Rating Scale (CRS)*	Observers
At the End of the Study	Final Teacher Questionnaire (FTQ)	Teachers
	Opportunity to Learn (OTL)*	Teachers
	Final Student Questionnaire (FSQ)	Students
	Posttest (POSTTEST)	Students

*Optional instruments

national researchers before sending their data to the international processing center at the University of South Carolina.

In order to assist in data cleaning and editing, the steering committee for the study agreed to a set of five rules for dealing with missing data. The rules provided a way to eliminate variables, students, teachers, and classrooms from a country's data set if the agreed-upon restrictions were not met. One of the rules, for example, stated that a teacher (and subsequently the teacher's class) would be eliminated within a country data set if the teacher had not completed both the Initial Teacher Questionnaire and the Final Teacher Questionnaire. See Anderson (in press) for information about the other rules.

Once the data were cleaned and edited, three data sets were created for each country: a student-level data set, a classroom-level data set, and a country-level data set. In creating the student-level data set, each student was of course assigned his or her own unique response (in a numerical code or score) on the variables from the four student instruments. For each variable from the four teacher instruments, each student was assigned a numerical value corresponding to his or her teacher's response. For each variable from the four observational instruments, each student was assigned the mean value (or alternative statistic, as appropriate) for the class over each observation period and across all observation periods.

The classroom-level data set was the same as the student-level data set except for the treatment of the variables from the student instruments. For these variables, the mean response or score (or alternative statistic, as appropriate) for all students in each classroom was used.

Finally, in order to examine the similarities and differences across the participating countries, country-level data sets were created by aggregating the data across all students, all teachers, and all classrooms within each country. These data sets included univariate statistics such as means and standard deviations for some variables, and frequencies and percents for others, as appropriate. The work described thus far was done by Garrett Mandeville at the University of South Carolina. When his tasks were completed, the data sets for each country were sent to Norbert Sellin at the University of Hamburg, where the multivariate analyses were to be performed.

In preparing the data for multivariate analysis, a careful examination of the raw data was made. Scatterplots and correlations of pretest scores with posttest scores were prepared

for each class in order to identify classes with peculiar pretest-posttest relationships. Classes in which the mean posttest score was lower than the mean pretest score by four or more points were noted. Classes with high nonresponse rates on specific instruments were identified. The frequency distributions of all variables were examined. Variables whose distributions were highly skewed were noted. Variables with little, if any, variance were identified. These peculiarities were considered when the correlations between the observational variables and student achievement were interpreted. Based on this careful examination and discussions among participants in this study concerning the results of this examination, classes and/or variables were eliminated from further analysis on a country-by-country basis. The results of this examination were included in the descriptive data pertaining to each country, however.

Once this preliminary work was completed, two multivariate analyses were performed. These analyses examined the nature and strength of the relationships among selected variables. Without going into detail (for which, see Anderson, in press, and the forthcoming international report), I will briefly describe the two analyses.

The first multivariate analysis used the student-level data sets and included 15 variables taken primarily from the student instruments. Thus, it was referred to as the "student variable model". The 15 variables were: home background, type of school, age of student, sex of student, pretest, aspirations of student for further education, attitude toward school, pre-attitude toward the subject, student engaged time, student perception of the task-orientation of the classroom, student perception of the extent to which the teacher provided structure for their learning, student perception of the extent to which the teacher provided feedback concerning the adequacy of the work and learning, opportunity to learn (which was included only in those countries which elected to use the OTL instrument), and the two outcome variables: student posttest and student post-attitude toward the subject. The student variable model was tested in six of the participating countries: Australia, Canada, Hungary, the Republic of Korea, the Netherlands, and Thailand. For Canada, the two Ontario replications (English and French) were both included: so there were seven replications tested. The remaining countries (and Quebec) were excluded because of small sample sizes, excessive amounts of missing data, or the apparent unreliability of the data pertaining to some of the key variables included in the model.

The model developed for the second multivariate analysis was tested on both the student-level and the classroom-level data sets. It was referred to as the "core model" for the study. Basically, this was an expanded version of the student variable model. In addition to

the 15 student model variables, six others were added from the teacher instruments and observation instruments. These six variables were: size of class, years of teaching experience, teacher perception of the students in the class (e.g. ability level, percentage of students needing remediation), direct teaching (e.g. providing cues, giving explanations, using examples), questioning practices (e.g. asking questions, reacting to student answers), and classroom management (e.g. engaging in procedural matters such as collecting homework, disciplining students for behaving improperly). These additional variables were selected in part because of preliminary analyses of their simple and partial correlations with student achievement and attitude. In this model, therefore, the set of teacher and classroom variables represented an empirically-based "best guess" of those classroom environment variables that may influence student achievement and attitude. Unfortunately, the core model could only be tested in four countries since it required a minimum of 40 classrooms (remaining after data cleaning and editing). The four countries were Australia, the Republic of Korea, the Netherlands, and Thailand.

In both models the relationships among the variables were depicted in the form of a causal model or path diagram. The statistical procedure Partial Least Squares (PLS) was used to analyze the data in terms of these models. See Anderson (in press) or the forthcoming international report for a discussion of PLS and for details of the series of initial analyses, reanalyses, and remodeling that were performed in order to develop "country unique" models. It is worth noting here, however, that "the path analysis applied to these models may be seen simply as an elaboration of the analyses employed in earlier process-product research in which fixed characteristics are partialled out of the relationships between teacher behavior measures and posttest achievement" (Mandeville, 1984, p. 156).

Major Findings and Generalizations Derived from the CES

Despite the care given to the preparation of the data and the choice of appropriate data analytic techniques, the results of the IEA study were somewhat disappointing. Descriptively, the data obtained from the observational instruments, especially the Five-Minute Interaction instrument, failed to differentiate between the teaching practices of the various countries. Of the 28 interaction categories on the FMI, for example, six accounted for between 60 and 77 percent of all interactions observed in virtually all participating countries. Almost one-half of the interaction categories were rarely, if ever, used by the observers in any country. Within the context of the correlational/experimental design, these 13 interaction behaviors (e.g. asks a higher-order question, checks the effectiveness of an

explanation) would not be included in the experimental treatment since they could not correlate with student achievement, given their low frequency and lack of variance.

Inferentially, the data obtained from the observational instruments failed to yield large, direct effects in terms of student achievement within the context of the core model. Path coefficients larger than 0.10 were rarely observed in any country. Even the zero-order correlations of the observational variables with achievement were small. Furthermore, many of the stronger correlations could be attributed primarily to irregularities in the distribution of the data on particular observation variables (e.g. the presence of a single "outlier" teacher, severe skewness).

A full presentation and discussion of the findings of the Classroom Environment Study will be given in the forthcoming book being prepared under the editorship of Ryan, Anderson, and Shapiro. The brief synopsis of some of the findings given above was intended to prepare the reader for six generalizations which we derived from our findings. Two of the generalizations were derived from the preliminary analyses of the data. One generalization was derived from the univariate statistics. The three remaining generalizations were derived from the multivariate analyses performed within the context of the student variable model and the core model. The presentation and discussion of these generalizations are quoted directly from Anderson (in press).

Generalization 1. Within countries, teachers differ greatly in what they teach relative to what is tested.

As a national option teachers were asked to indicate the extent to which their students had had an opportunity to learn the content included in each test item on the Student Posttest. Based on their responses, an opportunity to learn (OTL) index was derived for each teacher. This index represented the percent of items on the Student Posttest that the teachers believed their students had had an opportunity to learn during the conduct of the study. For example, an OTL index of 75 indicated that a teacher believed his or her students had an opportunity to learn 75 percent of the items on the Student Posttest.

Because the OTL instrument was a national option, OTL data were not available for all participating countries. As a consequence Mandeville (1984) developed a procedure for estimating OTL empirically. Mandeville's reasoning was as follows. If the students had had an opportunity to learn the content tested by a particular test item, then at least five percent more students should answer an item correctly on the posttest than did so on the

pretest. Accordingly, one could compute the percent of items on which the increase in the proportion of students answering each item correctly was more than five percent. This percent of items would represent an empirical estimate of OTL.

Both estimates of OTL have obvious sources of error associated with them. The empirical estimates of OTL are not independent of the posttest scores and are subject to the influence of any factor which might influence student learning relative to specific items (e.g. unusually apt students who are able to deduce answers to questions concerning content not taught, unusually good teachers who help students apply or generalize their acquired knowledge to new and different situations). On the other hand, the judgmental estimates of OTL are not independent of the feelings teachers have as they complete the OTL instrument. Specifically, teachers who believe their instruction is being evaluated might respond that, as expected, they teach everything tested to all students.

In view of these sources of error, then, the empirically derived estimate (set at an increase of one student in 20) may be viewed as an underestimate of the actual OTL. The judgmentally derived estimate (because of the pressure on teachers to give socially acceptable responses) may be viewed as an overestimate. As is frequently the case, therefore, the overall "best" estimate may lie somewhere twixt these two. In fact, the data with respect to these two estimates support the above contention. In general, the empirical estimate of OTL is approximately 10 to 15 percent lower than the judgmental estimate (in those countries in which both the judgmental and empirical estimate are available).

Independent of the estimate, however, the variation among the classrooms within countries in students' opportunity to learn is striking. Using the judgmental estimate, the ranges of OTL within four of the participating countries are 0 to 100, 8 to 100, 28 to 97, and 30 to 100.[1] (The means within these four countries are 79, 67, 75, and 87, respectively.) Using the empirical estimates, the ranges of OTL within four of the participating countries are 20 to 88, 29 to 97, 40 to 93, and 48 to 100. (Once again, the means for these four countries are 52, 63, 65, and 85 respectively.)

---

[1]Readers may be interested in the Ontario (English language schools) findings. For our sample of teachers, the OTL index based on their own reports ranged from 28 to 97 percent. That is, one teacher reported that only 28 percent of the test items related to content he/she had taught that year, whereas the test items were well matched to the instructional objectives of the second teacher. The mean was 75 percent, and the median was 80 percent. The test items were taken from the IEA mathematics test and had been judged valid for Ontario's Grade 8 mathematics curriculum.

Regardless of which estimate is used, the difference between the classrooms in which the students had the most OTL and the classrooms in which they had the least OTL is never less than 50 percent in any country, that is in certain classrooms students had an opportunity to learn the content related to two to three times as many test items as did students in other classrooms.

In virtually all participating countries, then, students in different classrooms differed greatly in the opportunity they had to demonstrate on the posttest the content they had acquired. The implication of the generalization is quite straightforward. To the extent that opportunity to learn influences student achievement, students in these different classrooms will perform differently on the achievement tests independent of the type or quality of the instruction they receive. As a consequence OTL becomes an important extraneous variable in our search for effective instructional behaviors and teaching practices.

Generalization 2. Teachers are inconsistent in the ways they behave from day to day.

When attempting to differentiate among teachers in terms of their practices, strategies, or styles, it would be nice if these teachers were consistent in their behavior across days, lessons, or observation periods. Such consistency would allow a more reliable classification of teachers into various typologies. The results of the CES suggest that such is not the case. Teachers were found to be highly inconsistent in the ways in which they behaved from day to day. Stated simply, teachers did and said different things on different days of observation. Furthermore, they set up different activities (e.g. lecture, written seatwork, oral practice) from one observation to the next.

The support for this generalization comes from a statistic called "eta square." Eta square represents the proportion of variance in teacher behaviors than can be attributed to differences between the teachers themselves. The smaller the eta square coefficient becomes, the more difficult it is to reliably differentiate between teachers in terms of their behaviors, and the more likely it is that differences in behavior can be attributed to differences in settings, situations, days of the week, and other similar factors. When the eta square is less than 0.50, the variance in teacher behavior that can be explained by differences between teachers is equal to the variance that can be attributed to all other sources or factors. When eta square is less than 0.50, the variance in teacher behavior that can be attributed to other factors (e.g. settings, situations, and days) is greater than the variance attributable to differences between the teachers themselves.

With these issues in mind, consider the findings of the CES. Consider, for example, the results pertaining to a single activity, namely lecture. In no country did the eta square coefficient for lecture exceed 0.50, that is in no country could the tendency to lecture be attributed primarily to differences between teachers. Rather, the tendency to lecture could be attributed primarily to other factors (such as the day of the observation -- perhaps teachers tended to lecture more on Mondays than on Fridays; the intended outcome of the lesson being observed -- perhaps teachers tended to lecture more when they taught facts than when they taught problem solving; and the organization of the classroom -- perhaps teachers tended to lecture more when they arranged their classroom in rows than when they formed circles of chairs for discussion groups). This same result (that is, eta square coefficients less than 0.50) was found for 11 of the 52 (or approximately 21 percent) of the variables included on the observation instruments.

With respect to the other variables included on the observation instruments, the eta square coefficient exceeded 0.50 in only one of the participating countries for 18 variables (or approximately 35 percent). Thus, for slightly more than one-half of the variables included on the observation instruments, the variability of teacher behaviors across observation occasions was noticeable in all or all but one of the participating countries. In fact, for only one variable, namely, teacher acknowledgement of the correctness of students' responses to questions, could it be said that fairly reliable differences existed between the teachers in the majority of the participating countries (with five countries having eta square coefficients greater than 0.50).

The finding that observed differences in teacher behavior cannot be reliably attributed to differences between the teachers themselves casts doubt on the likelihood that the results of the CES can be used to design simple experimental studies of the effects of teacher practices or styles on student achievement and attitudes. Rather, the results suggest that a host of contextual variables (e.g. differences in instructional objectives, differences in types of lessons being taught, differences in the classroom organization, differences in the days of the week) contribute more to differences in teacher behaviors than do characteristic differences between teachers (e.g. a preferred set of behaviors). In fact, the results suggest that teachers likely adapt their behaviors to relevant situational variables in such a way as to behave <u>appropriately</u>, rather than <u>consistently</u>. Thus, teacher training programs likely to be successful should have a dual purpose; namely, (1) to help teachers develop a repertoire of potentially useful pedagogical skills and behaviors, and (2) to help teachers develop the abilities to decide when to use which skills or behaviors in their classrooms.

Generalization 3. The nature of classroom teaching is quite similar in all countries.

At first blush this generalization seems to contradict the previous one. If individual teachers differ so greatly in the ways in which they behave on various occasions, how is it possible to talk about the similarity of teaching? This apparent paradox can be resolved in one of two ways. First, this generalization is based on data aggregated at the country level, that is this generalization is based on the mean percent of time spent in various activities or engaged in all behavioral interactions by all teachers over all observation periods. Second, although teachers tend to be inconsistent in the ways in which they behave over time, the repertoire of behaviors they exhibit with any frequency within their classrooms is quite small. The observed inconsistency is caused by differences in their use of a limited number of activities and behaviors.

To illustrate this latter point, consider the following example. The SNAP instrument included a variable labeled activity. Ten types or categories of activites were included on the instrument (e.g. lecture, discourse, oral practice, and written seatwork). In general, the results suggest that three primary types of activities occurred to a greater or lesser extent in the classrooms in all participating countries. Teachers talk "at" or "with" their students (from 35 to 82 percent of the time across the countries); students work on assignments at the desks or at laboratory tables (from 15 to 47 percent of the time); and teachers engage in a set of general classroom management activities, such as taking attendance or distributing and collecting papers (from one to 16 percent of the time).

Differences between the countries tended to lie within these three general activity categories. In some countries, teachers tended to talk "at" the students (i.e. lecture); in others, teachers tended to talk "with" the students (i.e. discourse). In some countries, the talk focused more on new material to be learned; in others, the talk focused on the review of previously learned material. In most countries, students spent a large amount of time completing assigned work at their seats, desks, or laboratory tables. In most countries these were written assignments, although in two of the countries, both of which used science as the subject matter, at least part of the work involved projects or experiments.

Results supporting this generalization also came from the data on the nature of the teacher-student interactions obtained from the FMI instrument. The FMI instrument contained 28 interaction categories (e.g. teacher uses an example, teacher probes, teacher asks a higher-order question, student gives extended response, absence of verbal interaction, social

interaction, procedural interaction). With the exception of one country, six of these interaction categories accounted for more than one-half of all interactions that occurred in the classrooms (with a range across these countries from 56 to 77 percent). In combination, these six categories support the results pertaining to activities presented earlier.

Two of the categories pertain to teacher explanations, and two to the asking and answering of questions (primarily factual questions). These four interaction categories are aligned with the general activity of the teachers talking "at" or "with" the students. One of the six frequently coded interaction categories dealt with the absence of verbal interactions, a condition often found when students are working on assignments. The final frequently coded interaction category was labeled "procedural interactions," the type of interaction seen while the teacher is conducting the "business" of the classroom, or using previous terminology, engaging in classroom management. While there were once again certain differences in the frequency of use of these different types or categories of interactions in the various countries, the similarities across the countries were striking.

Another way of stating this third generalization is to say that the nature of classroom teaching across the countries is more similar than it is different. This generalization was somewhat unexpected in view of the vast societal and cultural differences between the participating countries.

Generalization 4. The history of the learner exerts a potent influence on student achievement and attitudes.

Stated simply, the prior learning and attitudes students bring to particular classrooms and courses influence what they learn in those classrooms and from those courses and how they feel about the subject matter being taught. In general, this influence is so great that few if any variables can compensate for it, let alone overcome it.

While this generalization is certainly not new in the field of education, the results of the CES contribute to our increased understanding of the generalization in two ways. First, the history of the learner is far more complex than the specific knowledge, skills, or attitudes that are brought to the study of a particular subject at a particular point in time. The type of home in which the child is raised, the aspirations students have for their future education and formal schooling, and the attitudes students hold toward school in general all tended to influence student achievement and attitudes towards specific subject matter in virtually all participating countries. In general, the results suggest that both learning and failing to

learn are cumulative. Although the strength of the influence of prior history on present knowledge and feelings may be diminished somewhat by using different measures at the beginning and end of a course of study (as was true in a few countries), the influence of years of learning well or poorly is not likely to be overcome during a relatively short time period regardless of the intensity or power of the educational experiences encountered.

Second, the influence of the students' home background is primarily on the entry-level achievement and attitudes, not on the achievement and attitudes with which students leave the classroom or course of study. Stated differently, the home environment does not seem to directly effect what students learn during a particular time period. Rather, the home background seems to serve a preparatory function, not an actualization function. As a consequence, the effects of the home background on student achievement and attitudes could be minimized if ways could be found to intervene dramatically between the beginning and end of some course of study.

Generalization 5. Students' perceptions of their classroom and the instruction they receive influence their achievement and attitudes.

Student perceptions of the task-orientation of their classrooms, that is the extent to which they perceive the classroom as being businesslike with an academic focus, exert a weak, but consistent, influence on student achievement and attitudes across countries. In addition, student perceptions of the degree to which their teachers provide the necessary structure for their learning also exert weak, but consistent, influences on student achievement and attitudes. Providing structure implies that teachers emphasize the important ideas and main points, give clear explanations, make students think about solutions to problems, and periodically give summaries of major issues. The influence of these perceptions is somewhat stronger and more consistent on student attitudes than on student achievement.

At least two implications of this generalization can be drawn. First, student perceptions of what takes place in classrooms may be more important than what actually takes place. This implication will be supported to a greater degree by the final generalization. Second, the creation of classrooms in which a social norm for academic work and quality learning is established, and in which teachers provide the necessary structure in order for students to separate the "wheat from the chaff", seems to be a sensible beginning in the search for more effective classroom environments and more productive teaching practices.

Generalization 6. Observed classroom activities and teacher behaviors tend to exert virtually no influence on student achievement and attitudes.

As was mentioned earlier, two models were developed in preparation for the multivariate analyses. The first model included primarily student variables (e.g. home background, entering attitudes toward school, prior achievement relative to the subject matter being taught). The second model included these student variables and, in addition, a set of variables associated with the teacher questionnaires and observation instruments. One estimate of the influence of the observed classroom activities and teacher behaviors on student achievement and attitudes is possible by comparing the proportion of variance in student achievement accounted for by the second model with the proportion of variance accounted for by the first model. Such a comparison was possible in the four countries for which both models were examined (namely, Australia, the Republic of Korea, the Netherlands, and Thailand).

Across these four countries the student variable model accounted for between 34 and 56 percent of the variance in student achievement, and between 22 and 42 percent of the variance in student attitude. When the variables of class size, years of teaching experience, teachers' perceptions of their students' abilities, direct teaching, questioning, and classroom management were added to the student variable model (to create the core model), the increase in percent of variance accounted for by these six additional variables ranged from 0 to 2 percent for student achievement, and from 0 to 6 percent for student attitude.

Additional support for this generalization comes from the simple zero-order correlations, that is the relationship between two variables. For the composite of behaviors included in the variable "direct teaching," the zero-order correlations with student achievement across the four countries were -.03, -.02, .07 and .21. For the composite of behaviors included in the variable "questioning", the zero-order correlations with student achievement across the four countries were -.05, .01, .07, and .10. Finally, for the composite of behaviors included in the variable "classroom management", the zero-order correlations with student achievement across the four countries were -.06, -.13, -.17, and -.22. This last set of correlations, which are the strongest and most consistent, suggest that students achieve less well in classrooms in which more time is spent by the teacher engaged in behaviors related to classroom management.

This final generalization, while initially troubling given the second primary purpose of the CES, was hardly surprising in light of several of the other generalizations. Students in

different classes differed greatly in their opportunity to learn the content included on the final achievement test. As a consequence, what teachers and students said or did in their classrooms may or may not have been related to the content students needed to learn to do well on the achievement test. In addition, teachers varied greatly in their use of particular activities and behaviors from one observation to the next. Thus, the aggregation of the observational data across occasions likely introduced a great deal of error into the data. Finally, such error decreases the likelihood of finding large correlations between activities and behaviors, on the one hand, and student achievement and attitudes on the other.

Based on the aggregated data, teachers were quite similar in the percent of time they spent engaged in specific activities or exhibiting particular behaviors. More importantly, there were behaviors which the vast majority of teachers in most countries rarely exhibited (e.g., using examples, asking opinion questions, saying a student's statement or response was incorrect, giving the students answers to questions asked in class, checking for students' understanding of an explanation, stressing the importance of what was being discussed). These variables tended to have mean percents near zero with very small variance between the teachers. As a consequence, the likelihood that these variables would, in fact, correlate with student achievement or attitudes was virtually non-existent.

Two Secondary Analyses of CES Data

Additional analyses of CES data sets have been undertaken by various participants in the study. I will briefly describe two of these. The first was conducted with data from the Grade 8 mathematics classes included in the Ontario sample in English language schools.

Hanna and Ryan (1985) first identified classes for which the achievement gain over the school year was significantly higher than for other classes with similar students. For the total sample of 27 classrooms, pretest means ranged from 13.5 to 19.8 on a 40-item test; the posttest means ranged from 17.3 to 29.2; the average achievement gain over the school year ranged from 2.8 to 15.6. The students in the class of Teacher 331 had a mean pretest score of 13.6, the second lowest in our classroom sample. Their mean posttest score was 29.2, the highest of all the classes. Their average achievement gain was 15.6, and virtually every student in the class made significant gains across the year. The students of Teacher 101 had a mean posttest score of 16.7 and a mean posttest score of 26.6 -- an average of 9.9, which was the second highest in our sample. These two teachers (331 and 101) were our "high gain" teachers.

Matching teachers 331 and 101 with other teachers whose students had similar average pretest scores, we compared the "low gain" teachers and the "high gain" teachers with respect to a variety of student, class, and teacher characteristics. Our analyses indicated that the significant gains in mathematics achievement in the "high gain" classes could not be attributed to student or teacher background. For example, the students in Class 331 came from families with low socioeconomic status and, according to Teacher 331, virtually all of them entered Grade 8 needing remedial instruction. Teacher 331 taught four other subjects in addition to mathematics and was not a mathematics specialist. In no respects were either the students in Classes 331 and 101 or the two teachers significantly different from the students and teachers in the "matched" classes in ways that would have helped explain the high achievement gains.

From the aspect of teaching strategies, however, the analyses revealed several differences between the "high gain" teachers and their colleagues. These differences were not readily apparent in the data generated from the Five-Minute Interaction instrument, when those data were aggregated across occasions and lessons. The differences were apparent in teacher intentions for the observed lessons (from the Pre-Observation Interview), in observed activity patterns and teacher roles in classrooms (from the Snapshot), and in teacher reports of their "preferred" strategies and practices (from the Final Teacher Questionnaire and Teacher Interview). We found that the "high gain" teachers represent, in the case of Teacher 331, an extremely organized approach to teaching wherein material is taught until it is mastered, thereby reducing the need to review work frequently, and, in the case of Teacher 101, an approach which emphasizes the presentation of material followed by extensive practice in applying the material to new situations. Both teachers spent little time on classroom management and procedural matters. Both actively worked with individual students during seatwork assignments, which served as a means of assessing informally whether students understood the material. Consider this interview comment from Teacher 331:

> Mine are short-term objectives, always followed up to see if learning really took place -- although I teach mostly as a class, there is a lot of individual follow-up. I go up and down the rows while they are doing the work so they don't get a chance to become frustrated. That prevents discipline problems... I never repeat a unit; I don't test a unit until the students know the material. If I think they have not learned the material I will approach it in a different way. I use alternative strategies a lot.

Compare this approach with that of Teacher 101, which is different but seemed to be effective with his/her particular group of students:

> Once the topic has been presented and some examples have been shown, I expect them to work on their own. Instructing the class generally gives a lesson that may last the whole period and then they'll have a period to do the exercises. Generally I prepare lots of different examples for them to work through.

The two "high gain" teachers appeared to have adopted teaching styles for mathematics which differed in significant ways from those adopted by colleagues who taught equivalent subject matter to students with equivalent initial achievement levels. For example, the "low gain" teachers spent significant portions of time in lessons and lesson segments planned for review purposes; they spent less time in monitoring individual seatwork and more time in classroom management and procedural matters; they were also uninvolved with students for greater proportions of time. In contrast, in planning and teaching their lessons, the "high gain" teachers placed particular emphasis on new material rather than reviewing old material. While their own teaching practices were different, these differences may be a reflection of the success of both teachers in diagnosing the particular abilities of their students and adopting teaching strategies that were appropriate for those students.

In summary, the secondary analyses of data from one of the Ontario studies supported a hypothesis that teachers' planned intentions are more important than the average frequency of their use of particular interaction types in understanding student gains in achievement. The teachers' intentions were reflected in their content coverage and in the task-oriented classroom climates they created. Because they differed in the kinds of lessons they planned (in the way they used seatwork, for example), no one "right way to teach" was apparent. Both teachers were effective, in our opinion, because they planned appropriately for their students and they implemented their plans.

The second secondary analysis of CES data was conducted as a result of the international data analyses -- specifically, the aggregation of interaction behaviors across occasions, lessons, and teachers. The results of our initial analyses had led us to suspect that the aggregations were eliminating teacher intentions, rather than eliminating observer error. We also reasoned that tests of the relationships between lesson-level teacher behaviors and lesson-level student behaviors (either learning or student participation) would provide more meaningful and interpretable findings.

For these reasons, Mandeville (1984) used lessons (rather than classrooms) as his unit of analysis in a secondary analysis of the international data. He first tested the correlations between observed teacher behaviors (SNAP and FMI) and observed student engagement (taken as a lesson-level "proxy" for student achievement). He found, first, that there was

exceptional variation between teachers regarding both the direction and magnitude of these relationships when the lesson is the aggregation unit. For example, the correlation between the percent of time that one teacher spent lecturing and the percent of students engaged during this activity was negative (-0.65); across all lessons taught by this teacher, the highest rate of student engagement was achieved during a lesson in which the smallest amount of lecturing was observed. For another teacher in the same country, however, the correlation between the time spent in lecturing and the percent of students engaged was positive (+0.87), and the lesson in which this teacher was observed to spend the least amount of time in lecturing was the lesson in which the rate of student engagement was the lowest. In short, what seemed to "work" for one teacher in gaining student participation or engagement did not necessarily "work" for another.

Second, Mandeville (1984) found that, if "typical" teacher behaviors were constructed by using average correlations (across teachers within a sample), in many cases the results yielded appealing patterns. That is, for the "typical" teacher, student participation rates were higher in lessons with the following characteristics: lecturing (particularly with support materials) in a large group context; encouraging whole group participation; using a variety of questions and redirects when necessary; responding and encouraging student responses to questions; indicating whether student's answers to questions are correct or incorrect; and keeping classroom management activities during lessons to a minimum. These broad-based characteristics yielded relatively consistent correlations across countries with student academic engagement rates, although great variation existed among parallel correlations for individual teachers. However, Mandeville did find one category of teacher behavior which had consistent relationships with student engagement across lessons and teachers. This was the area of classroom management. The data clearly pointed out that students are less attentive to academic issues during lessons in which discipline and other management activities are common, regardless of the teacher or country.

Although the results of these "averaging" analyses were somewhat congruent with what other process-product researchers have found, Mandeville (1984) cautioned against generalizing from the findings for two reasons. First, the findings do not apply to all teachers (with the possible exception of the importance of keeping classroom management time to a minimum). Secondly, they cannot be viewed as prescriptive since no causality can be claimed. Mandeville concluded, "Potent other variables such as characteristics of the lessons may well be the causal forces behind many of the relationships which have been identified" (1984, p. 161).

Given our emergent hypotheses about the importance of teacher intentions and plans for lessons, Mandeville (1984) also devised a way to examine the nature of lessons. Using data from three countries, he undertook a descriptive analysis of the sequence of activities within each lesson (e.g. a three-activity lesson might have teacher lecturing, then seatwork, then more lecturing). He found a large number of different sequences of activities across teachers. In one country, for example, the 323 observed lessons were characterized by 158 unique sequences of activities.

For the Ontario (English) sample, Mandeville found that 17 percent of the 173 lessons contained but a single activity. Surprisingly, in one-third of these 30 single-activity lessons, students were assigned written seatwork for the full period. The remaining 20 single-activity lessons involved either lecture, or review, or discourse respectively. The most common lesson structure (28%) contained two segments. These lessons were of two types: (1) a discourse segment followed by a written seatwork assignment or (2) a lecture segment followed by written seatwork. Of the remaining 93 lessons, there were 54 unique structures. Most contained between two and four segments; the most complex lesson consisted of four review and three lecture segments in alternation and closed with written seatwork. Finally, it is interesting to note that, of the 173 lessons, 65 percent began with an initial segment (lecture or discourse) which lasted an average of 50 to 60 percent of the class period. Around 75 percent of the lessons ended with students working at their seats, and these assignments averaged about one-third of the class period.

Mandeville's analyses support a generalization that teacher behaviors, taken out of context and aggregated, are not only unstable but also are related differently to student engagement for different teachers. Both the Ontario reanalyses and those of Mandeville lend support to the view that the timing and appropriateness of teacher behavior, rather than its frequency, are the important determinants of its effectiveness. Finally, both analyses lend support to the findings about the importance of reducing the classroom time spent on management or procedural matters and of establishing a task-oriented classroom climate.

Concluding Comments

The description of the IEA Classroom Environment Study portrayed the care and rigor that characterized our efforts to use a process-product research model to generate knowledge of effective teacher behaviors. Despite our best efforts, the study failed to yield such knowledge. In combination, the results of the international analyses did not support a

decision to continue with the initial plans, which were to use the correlational study findings to develop teacher training programs and to conduct an experimental phase of the IEA study. Therefore, the Classroom Environment Study was terminated upon completion of the correlational phase.

Although all of the aims of the CES were not accomplished, the study did provide the basis for a greater understanding of what happens in classrooms throughout the world than was available previously. Anderson (in press) observed:

> Knowledge about differences in content covered, the variability of an individual teacher's behavior over time, and the reliance of teachers on a relatively few activities and behaviors to carry them through the school year is useful for those attempting to design new, more productive studies of classroom environments and teaching practices. Knowledge concerning the central role of the students (both their prior history and their perceptions of what actually transpires in the classroom) in their own learning suggests the need to include students in all conceptualizations of classroom teaching.

Our efforts to conduct secondary analyses of the data were somewhat more rewarding in terms of providing clues to effective teaching. We were limited, however, by the research "tools" used in the international study. For example, we began to suspect that teacher intentions are important keys to understanding effective teaching; yet we had little data about those intentions with which to test that hypothesis. We began to suspect that the classroom setting has powerful influences on a teacher's choice of interactive behaviors; yet we had little data to test that hypothesis. We began to suspect that the activities and events in classrooms are "mediated" in important ways by what students perceive and feel; yet, again, we had limited data to test this hypothesis. Each of these hypotheses, and the reasoning which led us to them, pointed to the need for a different conception of the teaching-learning process in classrooms. The process-product model for teacher effectiveness research falls far short of what we need to move from research findings to implications for improving teaching and learning.

III. A NEW MODEL OF TEACHER EFFECTIVENESS BASED ON A TASK VIEW OF TEACHING

My colleagues and I have thought long and hard about our shared experiences and frustrations in working, for four years, on the IEA Classroom Environment Study. In retrospect, the kinds of criticisms of process-product studies that were being aired in the literature by 1978 came back to haunt us. From our individual and collective reflections, one generalization emerged: the validity of the assumptions underlying the process-product model or paradigm must be questioned.

It is instructive to review the underlying assumptions of the process-product model and the criticisms of those assumptions that were present in the literature and which were borne out in our experiences in the IEA study. Anderson and I (1984) identified three basic assumptions that influenced both what we studied and how we proceeded in the IEA study.

Assumption 1. Recommendations for educational improvement can be empirically derived.

This assumption means that an a priori theoretical or conceptual framework is not needed as a basis for designing educational improvement efforts. Instead, researchers can simply examine what "works" under natural conditions in classrooms and use the results of these empirical investigations to design teacher-training materials.

Assumption 2. A linear, two-phase correlational/experimental design is appropriate for improvement-oriented research.

This assumption follows from the first. The correlational phase is needed to identify what "works" for "good" teachers; the experimental "treatment" (a teacher-training program) is derived from the naturalistic study; and the experimental phase is needed to test whether the "treatment" works with teachers and to test whether the "taught" behaviors result in improved student learning. This two-stage design is assumed to be a useful way to approach improvement-oriented research because, by definition, teachers will be taught to use the kinds of behaviors that "good" teachers already use. Teachers will not be asked to use

"esoteric" behaviors or practices that are "only theoretically good". Stated differently, as Gage and Fordham (1980, p.4) wrote in the proposal for the IEA CES, "It is easier to change a little than it is to change a lot about the school."

Assumption 3. The process-product paradigm is an appropriate paradigm for studying classroom teaching and learning.

Again, this assumption follows from the other two. Stated simply, the process-product paradigm assumes that teacher behaviors and practices (process) influence student achievement (product). Intuitively, it makes sense that improvement-oriented research should focus on what teachers do in the classroom and the relationship between what they do and what students learn. This is much more useful than focusing on teacher traits (personal or professional characteristics and qualifications of teachers). Moreover, this paradigm is an attractive alternative to the input-output paradigm because we all believe that teachers do make a difference.

Rethinking the Validity of the Process-Product Assumptions

Based upon our experiences in the IEA study, the three assumptions stated above, however appealing they may be, seem inappropriate and inadequate to guide improvement-oriented research. Our views are shared by other researchers. For example, Doyle (1978) has expressed strong reservations about the process-product paradigm specifically because of its underlying assumptions. He wrote that the paradigm contains few explicit explanatory principles to guide the selection of variables or the interpretation of results. Its two-factor structure ignores student learning behaviors and other variables that might contribute to explanations of how the teacher behaviors affect student learning outcomes. In practice, wrote Doyle, the selection of variables seems to be made on the basis of either personal preferences or on strictly empirical criteria such as the magnitude of correlation coefficients. Although the summary statements made by Rosenshine (1976) and others were useful, Doyle argued that such summaries did not explain why such teacher behaviors or variables should be expected to relate to achievement. "In the absence of formal explanatory propositions, it is difficult to interpret contradictory findings or select other potentially fruitful avenues of investigation (p. 167)."

In the case of the IEA study, we experienced these problems directly. Our long list of variables appeared to be a "laundry list" of various people's "favorite" behaviors. The

behaviors we studied were so numerous and represented such "small bits" of classroom life that we had no way to reconstruct, from them, the classroom as a total environment. Lacking any theoretical or conceptual framework for selecting the variables that went on our "laundry list", we were hard pressed to understand why we ended up with almost nothing from any of the teaching constructs that predicted student posttest achievement. The secondary analyses I mentioned at the end of Section II represented our efforts to try alternative ways to reconstruct classroom environments and examine processes that might influence student learning. As noted, however, we were constrained by the nature of the data yielded from our atheoretical design.

Another criticism of the process-product paradigm is its almost exclusive focus on teacher variables, and only selected dimensions of the teacher's work at that. Doyle (1978) criticized two of these dimensions specifically. The first of these dimensions, frequency of a behavior, was due to the use of a low-inference observation schedule. Such systems record discrete instances of specific behaviors. The implication is that the number of times a behavior occurs determines its effects--more is better. The assumption that frequency determines effects, however, appears to have little empirical or conceptual support. He cited the studies on Brophy and Evertson (1974) and Soar (1972) as having suggested that for some behaviors optimum level, rather than absolute quantity, determines effects on student learning. Doyle also noted that for some behaviors the timing of use is more important than the frequency of use. Certainly, Mandeville's (1984) analyses of IEA CES data support this criticism as well.

Process-product studies have also been criticized because of their focus on teacher behaviors that are stable and broadly applicable across setting, learner, and task conditions. If a variable has different effects at different times or with different groups of students, it is of little practical value for purposes of designing teacher-training programs. Taken to the extreme, then, the process-product researcher ends up looking for "laws" about teacher effectiveness that, in Doyle's (1978, p.169) view, are "context-free, teacher-proof, and even student-proof".

As the IEA study results demonstrate, teacher behaviors tend to be unstable (i.e. they vary greatly across lessons and occasions for individual teachers). The process-product paradigm, however, presupposes a high degree of stability of a teacher's behavior; otherwise, they could not relate the observations made during a short period of time to outcomes measured at the end of the study. Not only was this assumption disproved empirically in the IEA study, but it can also be questioned on the grounds of teacher intentions which lead teachers

to adapt their behaviors to meet the occasions that arise in the classroom. Medley (1982, p.1898-9), for example, concluded:

> If, for instance, a teacher deliberately asks more questions when introducing a new unit than when reviewing one, and encourages pupils to ask more questions during review than when a new unit is being introduced, this fact would not be detected because these variations would be averaged across both types of situations. It is obvious that intentional changes in teaching style may be important elements of teacher performance skill; the wisdom a teacher shows in making such changes may well be a major determiner of how effective the teacher is. This important element in teacher performance has been almost completely ignored in process-product studies.

Finally, the results of the IEA study suggest that teachers have a limited repertoire of behaviors that they use in the interactive classroom process. Rarely if ever, for example, did we observe teachers giving examples or checking to see if students understood an explanation. In a more recent study, Anderson (1986b), found that 88 secondary school teachers he studied rarely presented clearly stated instructional objectives at the beginning of a class period; they rarely communicated to their students that they were capable of learning and expected to learn the assigned material; they rarely provided students with incentives to stimulate learning efforts; and they rarely checked to see that students had completed or reviewed the assignments. The process-product paradigm would require that these rarely occurring behaviors be excluded from the teacher training programs derived from the naturalistic, correlational phase. Good, Grouws, and Ebmeier (1983) found that the mathematics teachers they studied rarely devoted class time to a developmental portion of the lesson, even though mathematics educators stress the importance of taking time to develop the concepts or ideas to be learned by students. Again, had they been content to let their "naturalistic" study dictate their experimental training efforts, teachers would not have been encouraged to plan lessons to include a developmental segment.

From our reflections about our own findings and those of others, Anderson and I (1984) offered three alternative assumptions that could guide improvement-oriented research on teacher effectiveness. These three assumptions are:

1. A theoretical or conceptual perspective or framework is necessary for understanding and ultimately improving the teaching-learning process.

2. The design of experimental studies should be based on both a conceptual framework and on empirical evidence from a variety of studies (i.e. experimental as well as correlational studies).

3. The process-product paradigm is an oversimplification of the world of classroom instruction and school learning.

We believe that a new paradigm or model is clearly needed to guide future research on teaching and, thus, to yield findings that will be both valid and useful to school practitioners and teacher educators. This model should be consistent with the complexity of classroom teaching and should be oriented toward both understanding and improving instruction. Stated differently, we need a conceptual framework that provides a basis for examining and understanding the intentions of teachers, why certain teacher behaviors or practices are appropriate at some times and not at others and, thus, why teachers behave differently (inconsistently) from one occasion to the next. We also need a conceptual framework that would allow us to make use of "within-teacher" variation instead of trying to get rid of this variation by "cloning" teachers through the use of "context-free" teacher training programs.

The Task View of Teaching

In a special issue of Evaluation in Education (Ryan and Anderson, 1984), my colleagues and I presented one such conceptual framework. Our framework or model is based on the assumption that it is what teachers accomplish, not what they are or what they do, that makes them differentially effective. In the papers included in this issue of the above journal and in subsequent papers (see Ryan, 1985 a and b; Anderson, 1985, 1986 a and b), we have taken the view that good teachers are able to accomplish a complex set of functions or tasks. In combination these tasks, when accomplished by teachers, result in appropriate student classroom behavior, active student involvement in learning and, ultimately, high levels of student learning.

Such a view began to gain supporters during the end of the 1970s and the beginning of the 1980s, and the supporters came from a variety of disciplines including sociology, psychology, and educational research and development (Anderson, 1986; see also Rosenshine, 1983; Brophy and Evertson, 1978; Brophy, 1981; and Good, Grouws, and Ebmeier, 1983). In combination, these researchers emphasized two major problems with research based on the "behavior" view of teacher effectiveness. The first of these problems is that teacher behaviors derive their meaning from the context in which they occur; thus, one must pay attention to the appropriateness of teacher behaviors and not simply their presence or absence, frequency, or duration. The second problem is that most of the things that teachers do in their classrooms are done for some purpose so that the same behaviors may be intended to achieve different ends at different times; thus, one must pay attention to

teacher plans and intentions and not merely the consistency (or inconsistency) of in-class behaviors.

The task view of teaching allows us to address these two major problems. A task is defined as a "discrete organized unit of work with a definite beginning and end, performed by an individual to accomplish the goals of a job" (Gael, 1983). Tasks allow us to understand why teachers do what they do or, stated differently, what functions are being served by their behaviors (c.f. Rosenshine, 1983; Brophy, 1981). For example, are teachers asking questions to gain the attention of particular students (a management task) or to check whether students understand some new content (a teaching task)? Tasks also allow us to examine the appropriateness of teacher behaviors by focusing attention on the sequence of teacher tasks during school terms, weeks, or days of instruction. In our model of teaching, we assume that certain tasks must be accomplished at certain times during the school year. For example, if the tasks of setting and clarifying rules and routines to govern student behavior are not completed during the first two or three weeks of the school year, teachers will have to spend more time during the year to establish and maintain order, at the expense of accomplishing instructional tasks.

Finally, the task view of teaching provides a useful (and, we think, valid) framework for efforts to improve instruction. For teacher supervisors, for example, it is much easier to judge the inappropriateness of teacher tasks than the inappropriateness of individual behaviors of teachers. For teachers, the completion of a particular task has an immediate payoff and, thus, is appealing. I will return to the implications of this model for practice later. First, I will describe the conceptualization of teacher tasks that my colleagues and I developed. The discussion is taken from Ryan (1985).

A Framework of the Major Tasks of Teachers

The starting point for the conceptualization of teaching that appears in Figure 1 was our assumption that the professional practice of teaching requires the completion of a large and diverse number of tasks, many of which must be completed outside the classroom. Consider this brief summary:

> In most developed countries, a teacher, perhaps an aide, and 20 to 40 students inhabit the classroom for 5-6 hours a day for 180 days. Over approximately a 40-week period, the primary job of teachers is to develop in students a variety of cognitive and affective goals. To accomplish these long-term goals, teachers must plan and organize the curriculum for the school year, choose or develop activities that

> will maintain student involvement, manage the social and instructional interactions in a smooth and orderly fashion, and take student differences into account in such a way as not to bore the faster students nor frustrate the slower ones. This must be accomplished in the face of community expectations, administrative and organizational constraints, rather substantial breaks in learning during the school year, interruptions in class schedules, and students who are occasionally less than willing to participate in classroom activities (Burns, 1984, p.96).

The literature on teacher planning and decision making helped us organize the varied and diverse tasks of teachers. We found that the job responsibilities of teachers could best be characterized in terms of decisions which vary in nature depending upon two factors: when decisions must be made (a timing dimension) and the situational demands that require decisions about different elements of the intended content and process (which we called structural units). We reasoned that decisions taken at different times and about different things were also made for particular durations of classroom instruction. These intended durations, moreover, seemed to correspond with the way in which a school year is organized and structured. Finally, we thought about the work of teachers appropriate to both different decision points and structural units as defining different types of teacher tasks. Tasks are goals to be achieved by teachers and sets of behaviors leading to the accomplishment of goals; thus, we conceived of tasks as linking teacher behavior with intention or purpose.

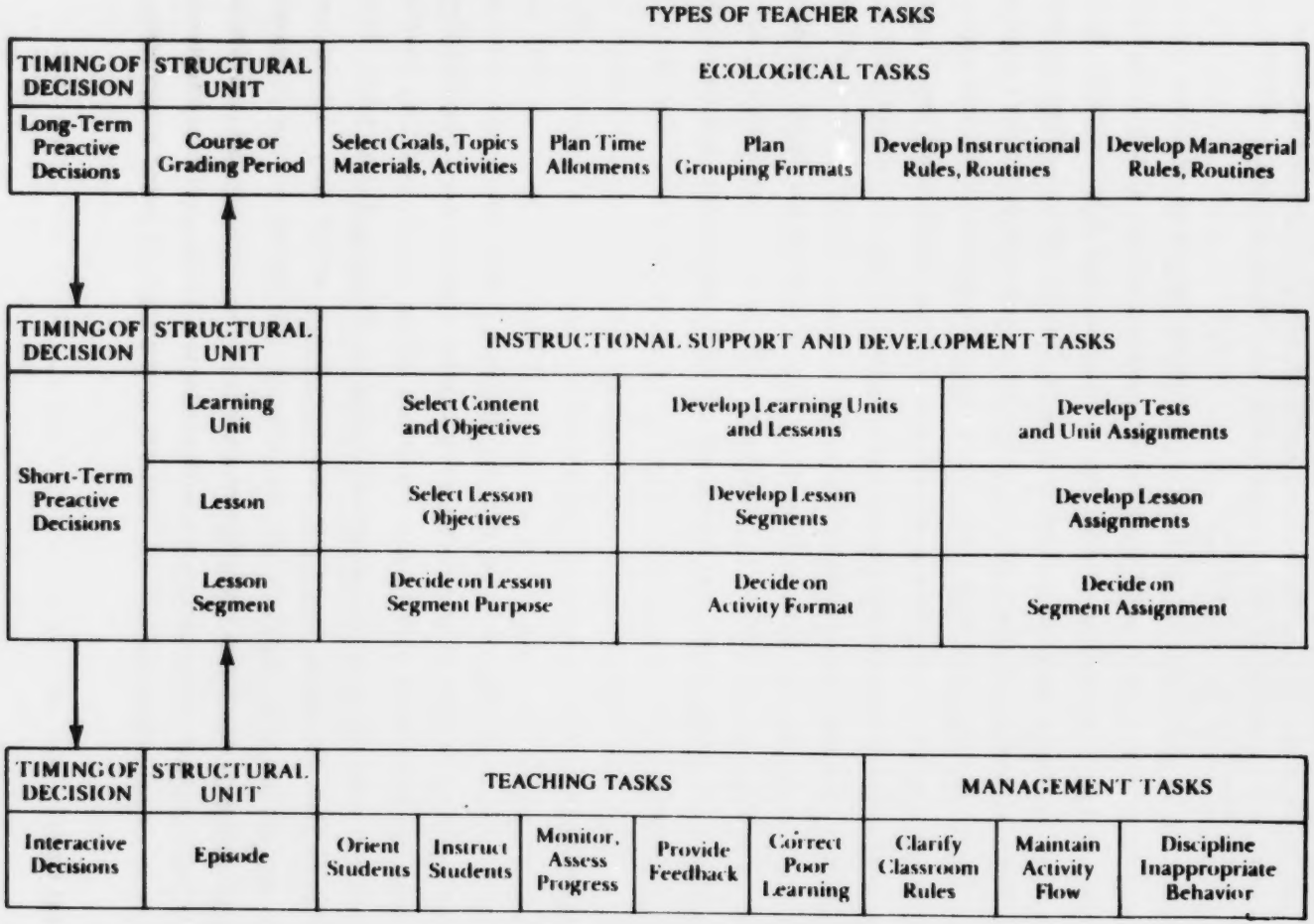
Timing of Decisions

The first dimension in Figure 1, timing of decisions, was based upon Jackson's (1962) distinction between preactive and interactive teacher decisions. Preactive decisions are taken while the teacher is outside the classroom (often alone), and they involve planning about what will take place in the classroom during a coming period of time. There are two kinds of preactive decisions, long-term and short-term, which vary in both specificity of planning and anticipated duration of the use of the plans in the classroom setting. Interactive decisions are taken while the teacher is in the classroom and involved in the flow of instructional events and interactions with students. These decisions involve on-the-spot judgments about behaviors and activity patterns as the planned lesson progresses.

Structural Units

Structural units, the second dimension in Figure 1, are the "things" about which teachers make decisions and, thus, are closely related to the timing dimension. In descending order by temporal boundaries, there are five structural units: course, learning unit, lesson, lesson

Figure 1. A Schematic for Conceptualizing Teaching in Terms of Decisions, Structural Units, and Teacher Tasks.



segment, and episode. Each lower unit is embedded within the next higher unit; for example, lessons are embedded within learning units.

The largest unit, course, is the natural temporal boundary within which school curriculum is organized, instructional activities are structured, and student learning is formally assessed. The course consists of learning units, the basic building block for organizing the curriculum. In making decisions about learning unit content, activities, and assessment, teachers generally plan for instruction across one of two weeks. Embedded within learning units are lessons, the basic planning and structural units of classrooms. In lesson planning, teachers prepare the objectives, activity sequences, and informal assessment to take place typically in one classroom period or day. Lesson segments, in turn, provide the activity format or structure and shape the expected roles of teachers and students for a certain number of minutes of lesson time. In mathematics classes, for example, teachers often plan a sequence of three segments for a lesson: review, development of a concept or procedure, and a seatwork assignment.

Embedded within lesson segments are episodes, which consist of bounded interaction or behavior sequences. There may be a large number of such episodes within each segment of the lesson. The concept of behavioral episodes was derived from the work of ecological psychologists (Barker, 1963; Gump, 1964) as representing the smallest meaningful unit of observable classroom behavior. Episodes are characterized both by a behavior or action and by the situational context (that is, the activity structure) within which the behavior occurs.

It is important to note that the meaning of each lower structural unit is defined by the next higher unit in which it is embedded. Thus, the meaning of a behavioral episode depends upon the activity context (segment) in which it occurs. A segment derives its meaning from the nature of the lesson in which it occurs. The meaning of a lesson must be derived from the learning unit of which it is but one component in a planned sequence of instruction.

Types of Teacher Tasks

The third dimension shown in Figure 1, types of teacher tasks, provides the means for identifying related units of work (tasks) to be performed in fulfilling the functional responsibilities of the job of school teacher. We identified four major types of teacher tasks, each of which is nested within decisions to be taken at different times during the school year and with regard to different elements (structural units) of the intended instructional content and process.

We labeled the first set of tasks ecological tasks since they serve the function of establishing broad curricular, temporal, and social constraints within which classroom instruction will occur. From Burns's (1984) review of ethnographic studies of teachers, we identified five sub-tasks to be performed early in the school year: selection of course goals, topics, materials, and activities; planning of time allotments for the topics; deciding if, when, and how students will be grouped for subject matter areas; development of instructional rules and routines; and development of management rules and routines. Each of these sub-tasks requires long-term preactive decisions related to the single structural unit of the course.

The long-term preactive decisions taken as teachers perform ecological tasks provide general guidelines for the subsequent planning for constituent elements of the course. It is important, therefore, to recognize that provincial or school system mandates may remove or constrain the teacher's responsibility for the first two sub-tasks (choosing curriculum content and materials and planning time allotments). Similarly, district or school organizational structures and policies (for example, class size, student placement procedures, course timetables) may limit the teacher's choice of grouping formats and compel the teacher to use particular instructional and management rules and procedures.

Given these general guidelines, teachers must perform a number of specific sub-tasks which, together, serve the function of providing the curricular and instructional basis for actual in-class teaching. We have called these instructional support and development tasks. Although the performance of each of these sub-tasks requires short-term preactive decisions, there are three different structural units involved: learning units (the constituent elements of the course), lessons (the constituent elements of learning units), and lesson segments (the constituent elements of lessons).

There are three instructional support and development sub-tasks to be performed for each of the three structural units. Thus, each row in this task domain in Figure 1 focuses attention on a distinct structural category for teacher planning. Each column in the task domain focuses attention on one of three important task elements in teacher planning: curricular goals or content, instructional activities and formats, and student assignments and tests.

Our delineation of instructional support and development tasks reflects the work of adaptive instruction researchers and developers and is congruent with a mastery learning approach to curricular and instructional planning (Anderson and Jones, 1981; Anderson and Block, 1976; Ryan and Schmidt, 1979). Mastery learning systems require teachers to attend

systematically to each of the short-term preactive tasks shown in Figure 1. For each learning unit, for example, a sequence of lessons must be developed in order to orient students to the intended objectives, to instruct students with regards to the content, to involve students in activities applying the content, to assess student learning, to provide feedback to students, and to correct student learning errors. An entire lesson might be developed for a single purpose (for example, to involve students in practice activities), but the important factor is that the lesson's purpose may be understood only in the context of plans for the unit as a whole. Furthermore, teachers may modify their timing plans for lessons in response to in-class events and student responses (for example, by providing an extra period for instruction or guided practice), but they would still be responsible for accomplishing all the sub-tasks intended for the sequence of lessons planned for the unit.

Teachers' functional responsibilities for implementing their short-term preactive decisions are defined by two interrelated sets of sub-tasks, teaching tasks and management tasks. Teaching tasks serve the function of involving the students in the planned lesson activities and learning tasks, and management tasks serve the function of maintaining both the flow of instruction and the involvement and cooperation of students in the planned lesson activities. Both teaching and management tasks require interactive decisions by teachers in response to minute-by-minute student actions and interactions. Both types of sub-tasks are performed with reference to the structural unit of behavioral episodes.

Again, Burns's (1984) literature review served as the basis for identifying sub-tasks required for the accomplishment of in-class teaching and management tasks. In his analysis of a number of large-scale process-product and experimental field studies and of reviews of these studies, Burns found there was strong consensus about a general pattern of effective teaching. This pattern or cluster of teacher behaviors is often referred to as direct instruction, and it consists of five interrelated teaching sub-tasks: orienting students, instructing students, monitoring and assessing the progress of students, providing feedback, and correcting poor learning when it occurs. There is a very close parallel between this cluster of effective teacher in-class tasks, identified from research on teaching, and the effective preactive tasks for planning a sequence of lessons within a learning unit, which was identified from the work of adaptive instruction developers.

The literature on classroom management helped us identify three sub-tasks, the accomplishment of which assists in providing a classroom atmosphere conducive of effective instruction. The sub-tasks are: clarifying rules and procedures, maintaining the activity flow of lessons, and promptly disciplining inappropriate behavior.

Although there was a consensus among reviewers of research on teaching about the sub-tasks which are accomplished by effective teachers, Burns (1984) found that there was much less empirical consistency with regard to particular teacher classroom behaviors that were more or less effective in accomplishing a sub-task. What seemed to be important was the occurrence of some teacher behavior that could accomplish the goal of a sub-task, rather than the occurrence of a particular behavior. For example, teachers should perform the sub-task of providing feedback to students. However, it is not clear whether it is more effective to give praise or criticism in performing this task. We suspect that the effectiveness of a particular task-relevant behavior depends upon the classroom context at that time.

The concept of functional equivalency, borrowed from Smith and Geoffrey (1968), was useful in understanding that different teacher behaviors may accomplish the same goal, depending on the particular context. Similarly, the same behavior may serve different goals, again depending upon the instructional context in which it occurs.

In summary, then, the job of school teacher carried with it the responsibilities for accomplishing tasks related to four major functional domains: ecological tasks, instructional support and development tasks, teaching tasks, and management tasks. In accomplishing the sub-tasks appropriate to the four domains, teachers must make decisions that vary in terms of timing (long-term preactive, short-term preactive, and interactive decisions) and in terms of structural referent (course, learning unit, lesson, lesson segment, and episode). Long-term preactive decisions (ecological tasks) influence the short-term preactive decisions (instructional support and development tasks) that in turn influence the performance of classroom teaching and management tasks. In implementing their plans, teachers must make interactive decisions intended to maintain both the flow of planned instructional activities and high student involvement. These classroom experiences provide feedback to teachers about their instructional support and development decisions.

The schema in Figure 1 helps to illustrate that in-class teacher behaviors cannot be understood or evaluated without knowledge of the broader plans for unit and lesson instruction. Moreover, if teachers have not accomplished the preactive planning tasks, there is no reason to expect that the appropriate teaching sub-tasks will be accomplished in the classroom.

Implications of the Task View of Teaching

One of the important conceptual limitations of the behavior view of teaching (i.e. it is what teachers do that makes them effective) is that the only reference point is student accomplishment of learning tasks. Medley's (1982) useful definitions which were included in Section 1 draw attention to this. He noted that, when teacher effectiveness is judged on the basis of student outcomes, any individual teacher might well be "effective" with one group of students and not with another. In the language of researchers, the effects of teachers on different groups of students are "unstable" and "inconsistent" from one year to the next. Rosenshine (1977) even found that these effects are unstable from one topic to another for the same group of students.

Obviously, all of us wish to find instructional strategies that, over the long term, result in consistent improvements in student learning. However, we believe it is important to think of the accomplishment of various teacher tasks as having more immediate outcomes for teachers. For example, when a teacher accomplishes the task of deciding on behavioral rules and routines for his or her classroom (an ecological task) and when he/she introduces these and clarifies them for students (a classroom management task), there are immediate and continuing "payoffs": student understanding of, and adherence to, those rules and routines; smooth transitions in classroom activities; and more time available for instruction. Similarly, when a teacher has accomplished an assessment task, the "outcome" or payoff is that the teacher gains an understanding of the current level of knowledge, skill, and/or motivation of the students. This understanding can be used to make a variety of important decisions (e.g. Should I group students for instruction? Should I move on to the next topic? Should I assign this student a failing grade?).

Endorsement of the view that it is what teachers accomplish that makes them effective allows us to see teaching as a series of interrelated tasks, each of which has specific, identifiable payoffs for the teacher. As Anderson (1985) notes, none of the payoffs of individual teacher tasks is student achievement. Rather, the task view of teaching assumes that "only when the complete set of tasks has been accomplished over a fairly extensive period of time can student achievement be expected to be affected substantially" (Anderson, 1985, p.21).

From the research conducted from a variety of perspectives over the past decade, we have available useful knowledge about teacher tasks and about how teachers may accomplish these tasks in effective ways. Broadly interpreted, the findings from ecological

psychologists, ethnographers, adaptive instruction developers, researchers on teaching, and sociologists point to the importance of: establishing a learning environment in the classroom that is conducive to student efforts to learn, organizing the curriculum, teaching to clearly thought-out objectives, and insuring that valuable class time is devoted to instruction and student efforts to learn.

What is needed in teacher preservice education is an examination of the curriculum and process from the standpoint of preparing teachers to accomplish the tasks that, together, define their job responsibilities and result in establishing classroom environments conducive to student learning. Conventional teacher education programs fall short of either serving or attaining these purposes. In another place (Ryan, 1984), I have provided a comprehensive review of conventional teacher education programs and inservice programs. Briefly summarized, I found, first, that theory and methods courses are separated from each other in such a way that teacher candidates have difficulty seeing how the theory courses can help them make decisions or "do their job". If these courses were conceived and taught with a focus on the types of teacher tasks for which they have appropriate content, the link would be evident. Moreover, knowledge from research would be presented with reference to its usefulness in helping teachers meet their every-day job responsibilities.

Secondly, I found important omissions in teacher preservice education programs with regard to several of the tasks that teachers must accomplish. In particular, there is little attention to the tasks that require different kinds of preactive planning by teachers. Theory courses, for example, rarely address the distinctiveness of the classroom as an ecological system (Doyle, 1977) or a management situation (Johnson and Brooks, 1979), so that prospective teachers gain little understanding of the grounds on which to make decisions about organizing a particular group of students for instruction and about choosing a set of viable rules and routines for student behavior (Calfee and Brown, 1979). Curricular planning is taught with a singular focus on individual classroom lessons (Romberg and Price, 1983), probably because teaching itself is narrowly defined as meeting with students (Goodlad, 1983). Little or no specific training is provided in developing or selecting appropriate instructional support materials (Courter and Ward, 1983) or in selecting or developing appropriate student assignments or tests or even in making decisions about the level of student performance teachers should expect (Stones, 1984; Ryan and Schmidt, 1979). Decision-making strategies are rarely taught (Borko et al., 1979).

In short, it appears that prospective teachers are assumed to learn what they need to do by being "apprenticed" to an experienced teacher. The focus is on the behaviors of this

"master" teacher in interactive classroom settings. This apprenticeship approach leads to attempts to copy "bits" of teaching behavior, the effectiveness of which may well depend upon being part of a total pattern of behavior: "when fragmented and adopted by another, they may be ineffective or even harmful" (Stones, 1984). Good, Grouws, and Ebmeier (1983) reinforced our argument that individual teaching behaviors do not constitute an appropriate way to conceptualize instructional behavior. Classroom behaviors derive their meaning from teacher intentions and preactive plans. The accomplishment by teachers of preactive planning tasks is much more important then the particular behaviors they exhibit in classrooms. The classroom behaviors are simply means by which plans are implemented.

The kinds of implications that the task view of teaching has for critically examining and reconceiving teacher education programs are paralleled by its implications for teacher supervision and evaluation. In another place (Ryan, 1985a), I have fully discussed these implications. Summarizing briefly, I found that current supervisory practices are grounded in a process orientation to instructional management, employ subjective judgments of teacher effectiveness based very largely on observations of the classroom behaviors of teachers, and result in prescriptions or directives about classroom behaviors ("hows").

Using the task conceptualization of the teacher's job (Figure 1), I argued that the focus of teacher supervision should be shifted to assistance and assessment relative to the appropriate accomplishment by teachers of preactive planning tasks ("whats"). The literature review that preceded our development of Figure 1 revealed that teachers rarely deviate from their plans, even when the lesson is going poorly (Peterson and Clark, 1978). Stated differently, teachers accomplish what they plan. Moreover, the very important planning tasks of teachers cannot be observed and, therefore, evaluated during classroom visits. Of five structural units requiring teacher task performance, only one is directly observable in classroom settings. One of the important "lessons" from research on teaching is that any individual classroom behavior may be effective or ineffective at any moment in time, depending upon <u>why</u> the teacher is using it and how appropriate it is for that purpose at that moment in that class. If teachers are accomplishing their planning tasks and if they are benefiting from the information about their students that is generated continually by their assessment tasks, the teachers are in a much better position than are supervisors to judge the appropriateness of their own behaviors. In essence, teachers are able to place these behaviors in context.

I also argued that teacher supervisors should not only stop thinking of teaching solely as meeting with students, but they should also stop thinking of teacher supervision solely as

observing and meeting with individual teachers. If the focus of instructional improvement in schools is shifted to helping teachers accomplish preactive planning tasks, teachers as a group could work together on these tasks, with organizational assistance (Howey and Vaughan, 1983). I have noted that individual teachers lack the time and skills needed to accomplish the necessary planning tasks. Thus, individual supervision should complement the major focus: school-wide or district-wide staff development programs and initiatives related to instructional support and development tasks.

Using Teacher Tasks to Link Research and Practice

The task or functional view of teaching provides an organizing model for examining the knowledge produced by research and for using that knowledge to improve practice. As noted above, considerations of discrete (but interrelated) teacher tasks provide a means of giving teachers a sense of accomplishment on an ongoing basis. Their effectiveness does not have to be measured only on the basis of student outcomes measured by tests. There is much to learn from the research conducted from a "behavior" view of teaching, but that knowledge is put in perspective, that is certain behaviors are probably more effective than others in helping teachers accomplish a particular task. And some teacher "traits" may be more likely to lead to task accomplishment than others (e.g. warm and caring individuals likely find it easier to accomplish the task of establishing rapport with students). The important point is that the "ends" of good teaching are seen as the accomplishment of teacher tasks; the traits and behaviors are likely to be "means" to those ends.

Placed in this perspective, I want to illustrate how knowledge from research can be helpful to teachers in attending to important tasks or functions. In working as a consultant to the State Education Department in South Carolina, Anderson (1986a) reduced Figure 1 into eight primary tasks and then described some of the things we know about how teachers accomplish these and what the "payoffs" are for teachers in accomplishing them. Because his article is not readily available to Canadian educators, I am reproducing much of his article here, in the hope that it will stimulate discussions among teachers, school administrators, and teacher educators. The next several paragraphs, then, are quoted directly from Anderson (1986a, pp.2-4).

Task 1. Planning Instruction

Instructional planning helps teachers achieve several purposes:

- To become aware of the important content and/or objectives, appropriate methods of assessing how well students have learned the content or mastered the objectives, and relevant materials and teaching methods.

- To ensure the availability of needed instructional support materials (e.g. supplementary readings, audiovisual aids, computer software), assignments, and tests.

- To estimate the amount of time to be allocated to various topics, objectives, book chapters, or activities (e.g. laboratory experiments).

- To align curriculum content and objectives, tests and assessments, and instructional methods and strategies. Teachers are able to direct their teaching toward what they expect their students to learn and assess student learning accordingly.

- To align the content and objectives of individual lessons or units with larger course goals. Students are able to see the relationship between isolated facts and the major concepts and generalizations that "define" the subject matter being studied.

- To design or establish instructional and managerial rules and routines.

Instructional planning occurs at a variety of levels. Not all teachers plan at all levels. Some plan at the course level and the daily level. Others plan at the weekly level only, relegating the course-level planning to a textbook and the daily planning to "whatever happens."

Novice and experienced teachers plan in very different ways. Novice teachers tend to prepare extremely detailed plans and follow them exactly. When their plans do not conform with reality, new teachers become confused about what to do. Experienced teachers, on the other hand, prepare rather sketchy plans that guide, but do not determine what happens in the classroom. If the plans "break down," teachers make "seat of the pants" decisions and improvise. Finally, teachers differ in their overall approach to planning (Clark and Yinger, 1979; 1980). Some are comprehensive planners who develop frameworks that guide their teaching based on their predictions of how things will (or will not) "work" in the classroom. Others are incremental planners, working in small steps, using day-to-day information to plan further action. Good teachers may differ in the way in which they plan and the form of their plans, but they all tend to accomplish the instructional planning task.

Task 2. Assessing Students

Teachers assess their students to gain an understanding of the level of their knowledge, skill, attitudes, or values. Airasian (1984) has identified three types of assessments frequently made by teachers: "Sizing up" assessments help them gain a general knowledge of students, individually and collectively, at the beginning of the school year or term. Instructional assessments show teachers how their students have benefited (or are likely to benefit) from

instruction as planned and delivered. Instructional assessments may include asking questions in class, giving students assignments, or administering teacher-made tests. Formal assessments are those required by law or state or district mandate.

Teachers can use assessment information to make various instructional decisions:

- To decide when to move students to new content or objectives.
- To assign grades to students or to decide which students should move to the next grade level and which should be retained.
- To make decisions about appropriate content and objectives for students. Such assessments focus on the extent to which students possess the necessary cognitive prerequisites for learning particular content or objectives.
- To determine which students (or subgroups of students) need extra assistance which will require additional effort on the part of the teacher, and arrangements for those students who do not need such assistance.

In any event, the form of the assessment is not as important as the information it provides.

Task 3. Clarifying Behavioral Rules and Routines

Clarifying behavioral rules and routines results in appropriate student behavior in the classroom. Students do not misbehave and react almost automatically when confronted with typical situations such as passing out papers, bidding for turns to talk in class, etc.

Rules and routines are so closely related that they are often confused. Simply stated, rules are statements of what is not permitted; they are sanctions or constraints on students' behavior in the classroom. Routines, on the other hand, are shared patterns of appropriate behavior. Routines provide acceptable and efficient ways of doing things that are done regularly in the classroom. While rules define inappropriate behavior, routines maintain the flow of activity in the classroom.

Several research studies have revealed at least three important characteristics of rules and routines that have become known and are widely accepted.

- Rules should be few in number and consistently enforceable.
- Important routines should be established early in the school year or as the situation arises.
- Rules and routines must be reinforced and maintained throughout the school year if they are to continue to be effective.

Task 4. Organizing the Classroom

Appropriate classroom organization can lead to two related payoffs:

- The vast majority of classroom time is spent on instruction and learning.
- The instructional format (e.g. lecture, recitation, media work, seatwork) and grouping arrangements (e.g. whole class, small groups, individualized) are supportive of the overall course aims and the level of student sophistication. The likelihood of students actually learning what they are expected to learn is enhanced.

"Within class" instructional grouping is a case in point. Individual differences in achievement among secondary school students are several times larger than the differences among primary school students. These differences even in so-called homogeneous secondary classrooms are likely to be larger than those in heterogeneous (or "non-grouped") primary school classrooms. Some within-class flexible grouping can solve the problem.

As expected, "within class" instructional grouping creates several problems for the classroom teacher -- problems that demand good classroom management. Fortunately, recent research suggests some solutions. Teachers must:

- Make clear assignments to those groups with whom they do not directly interact.
- Create in those groups what Kounin and Sherman (1979) refer to as "holding power" -- students tend to remain on-task even though the teacher is physically and psychologically removed from the group.
- Be aware of what Barr (1975) terms the "size of the remainder" -- the number of students not in the group with which the teacher is interacting. In general, the larger the size of the remainder, the more difficult the classroom management problems.
- Consider the feasibility and desirability of what has been termed "cooperative learning" (Johnson, Maruyama, et al., 1981), or "team-assisted learning" (Slavin and Karweit, 1981). In cooperative or team-assisted learning, a group of students work together to achieve a common task or goal.

Task 5. Creating a Learning Set

This task is known by several names, but "creating a learning set" (Hunter, 1983) is most descriptive. Creating a set results in students who are ready to learn. Readiness here implies that students have both the cognitive and affective prerequisites for learning.

- Students are aware of the specific content and/or objectives they are to learn.
- The association of the new content and/or objectives with past, present, or future learning is clear.
- Students are emotionally ready to learn. They have a true desire to learn or can be coerced into the effort needed to learn.

The task of creating a learning set is especially important for teachers faced with unprepared or reluctant learners. Recent research has shown that teachers can create a set by (1) providing students with the necessary cognitive prerequisites just prior to their teaching of new content and/or objectives, (2) clearly communicating their expectations about what students are to learn and how they are to demonstrate that learning, and (3) making use of incentives, reinforcement, and other principles of behavioral psychology.

Task 6. Teaching to Objectives

When this task is accomplished, students learn precisely what they were expected to learn. Teachers who teach to objectives have some instructional purpose in mind when they tell students what they tell them, show students what they show them, or assign the materials they assign them. Such teachers do not assign Macbeth simply because all students should read this classic; they assign it because they hope the students will learn about the development of plot or the moral struggle of the characters. Teaching to objectives includes both initial teaching and reteaching as necessary. In fact, the importance of reteaching is crystallized within the task of teaching to objectives.

Direct teaching and the clarity of expression are the principal techniques of teaching to objectives. Direct teaching occurs when the teacher is actively involved in transmitting clearly defined knowledge or skills to the students. Clarity of presentation implies that teacher transmission of knowledge or skills is made in such a way that the vast majority of students understand it and learn from it. Rosenshine (1983) offers several empirically-derived suggestions for helping teachers teach to objectives. Teachers should:

- Model or demonstrate the desired learning for their students (when appropriate).
- Organize and present one point at a time.
- Give detailed and redundant explanations for different points.
- Use many, varied, and specific examples to illustrate the major points.
- Check for student understanding before moving from one major point to another, asking probing questions when appropriate and necessary.
- Stay with the major points until all or almost all students understand them.

Task 7. Providing for Student Practice

Practice has two immediate payoffs. Students remember things more quickly and do them faster, and they remember what they have learned for longer periods of time. Practice

exercises and activities also serve an assessment function. Teachers find out how well students benefited from the initial teaching and which students may need additional instruction.

Two types of practice have been identified by researchers. Guided practice occurs under the supervision of the teacher, aide, tutor, or classmate, is typically of short duration, and is intended to help students succeed. Independent practice takes place away from the direct supervision of the teacher, aide, tutor, or classmate. Homework is a prime example of independent practice. The relationship between the two types of practice is quite clear: students should be helped to achieve a reasonable success rate in guided practice before they are permitted to practice on their own. Rosenshine (1983) again makes several research-based recommendations for the improvement of guided practice:

- Students should experience a high rate of success.
- "Prompts" should be used as necessary to lead students to correct answers.
- Students should be asked to explain their answers (e.g., "How did you arrive at that answer?").
- Errors should be corrected as they occur and not be allowed to accumulate.
- If common errors are made by a fairly large number of students, reteaching may be necessary.

Rosenshine also offers three suggestions about independent practice: directions for independent practice should be clear and concise; students' understanding of the directions should be checked before they are allowed to proceed; and to the extent possible, teachers should actively monitor practice and provide guidance as needed.

Task 8. Maintaining Student Involvement in Learning/Disciplining Inappropriate Behavior

Although it may appear that two separate elements are combined to form the eighth teacher task, they are in fact two sides of the same coin. The primary task is that of maintaining student involvement in learning -- keeping students engaged in learning or "on-task." The secondary task, disciplining inappropriate behavior, becomes increasingly important when students are not involved in learning. Students put forth the mental effort to learn well what they are expected to learn.

Anderson (1981, 1984) has identified a series of instructional factors that are associated with higher levels of student engaged time, or time-on-task. Students tend to be more involved in learning and spend more of their time on-task when:

- The assigned tasks are of appropriate difficulty.
- The assigned tasks are interesting or enjoyable.
- Performance on the assigned tasks is frequently monitored and assessed.
- The mechanical details of classroom behavior are reduced to a minimum.
- The physical conditions under which learning occurs are conducive to learning.
- Teachers maintain the flow of activity in the classroom.
- Student curiosity is aroused.
- Clear expectations are communicated to students.
- New learning is related to previous learning.
- Attention is focused on the relevant, important aspects of the instructional materials and activities.
- Feedback is provided on the adequacy or excellence of student task performance.
- Task-oriented behavior is reinforced. (Anderson, 1986a)

Others researchers have also tried to organize the knowledge from research into useful implications for practice. Rosenshine (1983) identified the instructional functions to which the findings of research could be applied. Stallings (1985) chose to organize research implications on the basis of attributes of effective classrooms and what teachers can do to create learning environments conducive to student learning. Stallings also identified available programs and materials for assisting teachers to develop a repertoire of appropriate skills and competencies. In the hope of stimulating interest to read Stallings's paper, I will include one illustration here. This is what Stallings (1985, p.166) says about classroom management:

> There is no doubt that students in classrooms that are well managed perform better on achievement tests (Brophy, 1981; Fisher et al., 1980; Good, 1979; Rosenshine and Berliner, 1978). "Because successful classroom managers maximize the time their students spend engaged in academic activities, their students have more opportunities to learn and this shows up in superior performance on achievement tests" (Brophy, 1982). However, knowing this fact will not help the preservice teacher know how to do it. Observations by researchers Evertson, Anderson, and Emmer in both elementary and junior high schools (1980) were so specific that practice can be guided even for the first days of school. These researchers describe in detail how effective teachers established and carried out their management plans, and subsequently developed a set of checklists for teacher use. Preservice teachers should have the benefit of reviewing these materials.

In short, the concept of teacher tasks or teacher functions provides a useful organizing model for thinking about teacher effectiveness and the findings of the large body of research on that topic. Reflecting upon his own attempt to use functions as an organizing heuristic, Rosenshine (1983) commented that he "looked at my old articles to see where sections fitted into the functions, and I was shocked to see how fragmented those articles were and the areas I missed." He concluded that a list of teacher functions (or tasks, in our terminology) is an excellent advance organizer which should be presented to teachers before they receive specific training programs.

Summary and Concluding Comments

In this monograph, I have reviewed briefly the kinds of research on teacher effectiveness that have occupied researchers over a period of more than eighty years. With each generation of researchers, the studies have built upon earlier findings and have attempted to address methodological problems and inadequacies of theoretical foundations. The most pervasive problem for researchers has been, and continues to be, establishing a starting point — a way to view teaching -- that could validly reflect the complexities of the job of school teacher. Only when research is designed from a valid starting point will the results offer useful directions for improving the performance of teachers.

Throughout the greater part of this century, research on teaching was guided by a general assumption that it is what teachers <u>are</u> that makes them effective or ineffective. Specifically, the early research was based on an assumption that teachers differed in their possession of certain personality traits that are characteristic of effective teachers. Thus, the research was focused on identifying the "desirable" traits, largely through collecting the judgments of former students and educators, and then incorporating the "desirable" traits into rating forms. The rating forms could then be used to judge the likely effectiveness of teacher candidates and practicing teachers. Both the assumptions about "good" teachers (i.e. stable traits) and the observation "tools" (i.e. rating scales) that were evident in the research of this era were adopted by school administrators and teacher educators, even though there was no evidence that teachers who possessed such traits were also those whose students learned more. The attempts to demonstrate such relationships (through presage-product studies) were largely unsuccessful and yielded inconsistent findings.

Starting in the mid-1960s, a new approach to research on teaching appeared and was labeled the "process-product" approach. The underlying assumption was that it is what teachers <u>do</u> that makes them effective or ineffective. Specifically, this research was based on an

assumption that teachers differed in their use of instructional behaviors that are characteristic of effective teachers. Thus, the research was focused on identifying the "desirable" teacher behaviors, largely through low-inference observation systems and through analyses of the relationships between the average frequency of use of various teacher behaviors and average student learning. The attempts to identify "stable" teacher behavior patterns and to demonstrate that these behavior patterns were related to student learning were by far the most successful in the history of research in teacher effectiveness. The findings of the process-product studies of the 1970s were incorporated into "teacher should" statements and inservice training programs. Behavioral checklists and classroom observations became commonplace in many school districts.

It was in this context that the IEA Classroom Environment Study was designed as a process-product study to be replicated in several countries around the world. Even as the international study was being designed, however, criticisms of the process-product approach were beginning to appear in the literature. The criticisms were addressed both to the assumptions being made about the process of classroom teaching and learning and to the methodology being used in these studies. It is fair to say that the researchers who cooperated in planning and conducting the IEA study were more influenced by the publications proclaiming the usefulness of such research than they were influenced by published criticisms of the approach. Our plans went forward accordingly, and we were to learn about the shortcomings of the process-product approach through our own direct experiences.

As a result of our work on the IEA study and our careful review of the literature, my colleagues and I began to question the validity of the "behavior" view of teacher effectiveness and, thus, of the process-product research paradigm. We suggested that a more valid and useful model of teacher effectiveness should be predicated on a "task" view of teaching. When teacher effectiveness is conceived of as the accomplishment of a set of interrelated tasks, there is a way to identify both serious omissions in teacher preparation and important errors in the way we supervise teachers and try to help them improve their instructional practices. The discussions in this final section of the monograph presented one such delineation of teacher tasks and the implications from that view for teachers, school administrators, and teacher educators. The concept of teacher tasks also provides an organizing framework to link research findings to implications for practice. This has also been discussed in Section III.

Obviously, the task view of teaching requires changes in the design and conduct of future research on teacher effectiveness. For example, attention to the preactive planning decisions that teachers make requires new methods of collecting information. One cannot directly observe whether, and how, teachers accomplish these tasks by visiting classrooms since these tasks require teacher attention outside of class time. Attention to the accomplishment of in-class teaching and management tasks also requires new methods of classroom observation. An observation system would need to be event-driven rather than time-driven, and the behavior categories included on them would need to be organized with reference to the relevant tasks (e.g. orienting, instructing, assessing, correcting, maintaining, and disciplining). In this regard, Anderson (1986b) has developed a structural observation system called DePICT (Describing Practices and Intentions of Classroom Teachers) which is being used both as a research tool and as part of the teacher evaluation process in the state of South Carolina. A full discussion of the implications for future research is available in the articles by Anderson (1984; 1986a and b); Ryan (1985b); and Mandeville (1984).

In conclusion, as we look to the future, the optimism dating from the late 1970s about research on teacher effectiveness should continue. We already have available, through a task or functional view of teaching, what Rosenshine (1983) calls "abundant riches for improving teaching practice". Moreover, the trend of enrolment declines in our schools seems to be on the verge of reversing itself so that school systems will likely be hiring large numbers of new teachers over the next decade (see Bridges, 1986). These opportunities could also result in improvements in the quality of teaching that will be experienced by future generations of students <u>if</u> we begin immediately to improve the education of teacher candidates.

Teacher education, both at the preservice level and the inservice level, should adopt as primary goals the development of competencies needed by teachers to create and maintain a task-oriented learning environment, to engage pupils in learning-related activities, and to use the kind of instruction that research indicates is provided by effective teachers. I agree with Medley (1979, p.26), who wrote:

> Nobody really feels that the effective teachers presently in schools....are effective enough; it is the state of the art of teaching itself that is in need of improvement. While the most urgent business of education is to improve the least effective teachers, improving the most effective ones is also important.

My colleagues and I believe that the creation of a task-oriented classroom climate and the productive use of learning time in classrooms are enabling conditions for effective learning by students. We have tried to suggest some of the major tasks that, if accomplished by teachers, should be of assistance in establishing such enabling conditions and, thereby, in fostering greater student effort and learning.

REFERENCES

Airasian, P.W. Classroom Assessment and Educational Improvement. Portland, Oregon: Northwest Regional Educational Laboratory, 1984.

American Educational Research Association. Second Report of the Committee on Criteria of Teacher Effectiveness. Journal of Educational Research 46 (1953): 641-658.

Anderson, H. A Study of Certain Criteria of Teaching Effectiveness. Journal of Experimental Education 23 (1954): 41-71.

Anderson, L.M.; Evertson, C.M. & Brophy, J.E. An Experimental Study of Effective Teaching in First-Grade Reading Groups. Elementary School Journal 79 (1979): 193-223.

Anderson, L.W. The IEA Classroom Environment Study: Teaching for Learning. Comparative Education Review (in press).

Anderson, L.W. Research On Teaching and Educational Effectiveness. NASSP Curriculum Report 15 (1986a): 1-4.

Anderson, L.W. Understanding Teacher Behavior in the Classroom: A Must for Sound Evaluation. NASSP Bulletin (May 1986b): 42-50.

Anderson, L.W. Thoughts on Accomplished Teachers. The Palmetto Principal (Nov. 1985): 19-21.

Anderson, L.W. Attention, Tasks, and Time. In L.W. Anderson (Ed.), Time and School Learning: Theory, Research, and Practice. London: Croom Helm, 1984.

Anderson, L.W. Instruction and Time-On-Task: A Review. Journal of Curriculum Studies 13 (1981): 289-303.

Anderson, L.W. & Block, J.H. Mastery Learning. In D. Treffinger, J. Davis & R. Ripple (Eds.), Handbook on Teaching Educational Psychology. New York: Academic Press, 1976.

Anderson, L.W. & Jones, B. Designing Instructional Strategies Which Facilitate Learning for Mastery. Educational Psychologist 16 (1981): 121-138.

Averch, H.A.; Carroll, S.J.; Donaldson, T.S.; Kiesling, H.J. & Pincus, J. How Effective Is Schooling: Critical Review and Synthesis of Research Findings. Santa Monica, Calif.: Rand, 1972.

Barker, R. The Stream of Behavior As an Empirical Problem. In R. Barker (Ed.), The Stream of Behavior. New York: Appleton-Century Crofts, 1963.

Barr, A.S. The Validity of Certain Instruments Employed in the Measurement of Teaching Ability. In H.M. Walker (Ed.), The Measurement of Teaching Efficiency. New York: McMillan, 1935.

Barr, A.S. & Emans, L.M. What Qualities Are Prerequisite to Success in Teaching? Nation's Schools 6 (1930): 60-64.

Barr, R. How Children Are Taught to Read: Grouping and Pacing. School Review 83 (1975): 479-498.

Barr, R. & Dreeben, R. Instruction in Classrooms. In L.S. Shulman (Ed.), Review of Research in Education (Fifth Edition). Itasca, Ill.: F.E. Peacock, 1978.

Borko, H.; Cone, R.; Russo, N.A. & Shavelson, R.J. Teachers' Decision Making. In P. Peterson & H. Walberg (Eds.), Research on Teaching. Berkeley, Calif.: McCutchan, 1979.

Boyce, A.C. Methods for Measuring Teachers' Efficiency. In the Fourteenth Yearbook of the National Society for the Study of Education, Part II. Bloomington, Ill.: Public School Publishing Co., 1915.

Bridges, E.M. The Incompetent Teacher. Philadelphia: The Falmer Press, 1986.

Brophy, J. Classroom Organization and Management. Paper presented to the Conference on Research on Teaching. Warrenton, Va., 1982.

Brophy, J. Teacher Praise: A Functional Analysis. Review of Educational Research (1981): 5-32.

Brophy, J. & Evertson, C. Context Variables in Teaching. Educational Psychologist 12 (1978): 310-316.

Brophy, J.E. & Evertson, C.M. Process-Product Correlations in the Texas Teacher Effectiveness Study: Final Report. (Research Report 74-4). Austin: University of Texas, Research and Development Center for Teacher Education, 1974.

Burns, R.B. The Process and Content of Teaching: A Conceptual Framework. Evaluation in Education: An International Review Series 8 (1984): 95-112.

Calfee, R. & Brown, R. Grouping Students for Instruction. In D.L. Duke (Ed.), Classroom Management, Seventy-eighth Yearbook of the National Society for the Study of Education, Part II, Chicago: University of Chicago Press, 1979.

Charters, W.W. & Waples, D. The Commonwealth Teacher Training Study. Chicago: University of Chicago Press, 1929.

Clark, C. & Yinger, R. The Hidden World of Teaching: Implications of Research on Teacher Planning. Research Series No.77, The Institute for Research on Teaching, Michigan State University, 1980.

Clark, C. & Yinger, R. Teachers' Thinking. In P. Peterson & H. Walberg (Eds.), Research on Teaching. Berkeley, Calif.: McCutchan, 1979.

Coleman, J.S.; Campbell, E.Q.; Hobson, C.J.; McPartland, J.; Mood, A.M.; Weinfeld, F.D. & York, R.L. Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

Comber, L.C. & Keeves, J.P. Science Education in Nineteen Countries: An Empirical Study, International Studies in Evaluation, Vol.1. New York: John Wiley, 1973.

Courter, R.L. & Ward, B.A. Staff Development for School Improvement. In G.A. Griffin (Ed.), Staff Development, Eighty-second Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago, Press, 1983.

Crawford, J. & Gage, N.L. Development of a Research-Based Teacher Education Program. California Journal of Teacher Education 4 (1977): 105-123.

Crawford, J.; Gage, N.L.; Corno, L.; Stayrook, N.; Mitman, A.; Schunk, D. & Stallings, J. An Experiment on Teacher Effectiveness and Parent-Assisted Instruction in the Third Grade. Stanford, Calif.: Center for Educational Research at Stanford, 1978.

Doyle, W. Paradigms for Research on Teacher Effectiveness. In L.S. Shulman (Ed.), Review of Research in Education (5th Edition). Itasca, Ill.: F.E. Peacock, 1978.

Doyle, W. Learning the Classroom Environment: An Ecological Analysis. Journal of Teacher Education 28 (1977): 51-55.

Elam, S.M. A Decade of Gallup Polls of Attitudes Toward Education 1969-1978. Bloomington, Indiana: Phi Delta Kappa, 1978.

Evertson, C.; Anderson, L.M. & Emmer, E. Effective Management at the Beginning of the School Year. Elementary School Journal 80 (1980): 219-231.

Evertson, C.; Anderson, L.M. & Emmer, E. Effective Management at the Beginning of the School Year in Junior High Classes. Austin: University of Texas, Research and Development Center for Teacher Education, 1980.

Fisher, C.W.; Berliner, D.C.; Filby, N.N.; Marliave, R.; Cohen, L.S. & Dishaw, M.M. Teaching Behaviors, Academic Learning Time and Student Achievement: An Overview. In C. Denham & A. Lieberman (Eds.), Time to Learn. Washington, D.C.: National Institute of Education, 1980.

Fordham, A. The Context of Teaching and Learning. Hawthorn, Victoria: Australian Council for Educational Research, 1983.

Gael, S. Job Analysis: A Guide to Assessing Work Activities. San Francisco: Jossey-Bass, 1983.

Gage, N.L. The Scientific Basis of the Art of Teaching. New York: Teachers College Press, 1978.

Gage, N.L. Research on Cognitive Aspects of Teaching. In The Way Teaching Is. Washington, D.C.: Association for Supervision and Curriculum Development and NEA Center for the Study of Instruction, 1966.

Gage, N.L. (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Gage, N.L. Address appearing in "Proceedings," Research Resume, Vol. 16. Burlingame, Calif.: California Teachers Association, 1960.

Gage, N.L. & Fordham, A. A Proposal for an IEA Classroom Environment Study: Teaching for Learning. IEA Secretariat (University of Stockholm), 1980.

Good, T. Teacher Effectiveness in the Elementary School: What We Know About It Now. Journal of Teacher Education 6 (1979): 105-113.

Good, T.; Biddle, B. & Brophy, J. Teachers Make a Difference. New York: Holt, Rinehart & Winston, 1975.

Good, T.L.; Grouws, D.A. & Ebmeier, H. Active Mathematics Teaching. New York: Longman, 1983.

Good, T.L. & Grouws, D. Missouri Mathematics Effectiveness Project: Experimental Study in Fourth Grade Classrooms. Journal of Educational Psychology 71 (1979): 355-362.

Goodlad, J.I. A Place Called School. New York: McGraw-Hill, 1983.

Gump, P. Environmental Guidance of the Classroom Behavioral System. In B. Biddle & W. Ellena (Eds.), Contemporary Research on Teacher Effectiveness. New York: Holt, Rinehart & Winston, 1964.

Hanna, G. & Ryan, D.W. Profiles of Effective Teachers of Grade 8 Mathematics. Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1985.

Hart, J.W. Teachers and Teaching. New York: Macmillan, 1936.

Hathaway, J. & Lockheed, M.E. Introduction: The Social Sciences and Education. In J. Hathaway & M.C. Lockheed (Eds.), The Contributions of the Social Sciences to Educational Policy and Practice: 1965-1985. Berkeley, Calif.: McCutchan, 1986.

Howey, K.R. & Vaughan, J.C. Current Patterns of Staff Development. In G.A. Griffin (Ed.), Staff Development, Eighty-second Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1983.

Hunter, M. Mastery Teaching. El Segundo, Calif.: TIP Publications, 1983.

Johnson, D.; Maruyama, G.; Johnson, R.; Nelson, D. & Skon, L. Effects of Cooperative, Competitive and Individualistic Goal Structures on Achievement: A Meta-Analysis. Psychological Bulletin 89 (1981): 47-62.

Johnson, M. & Brooks, H. Conceptualizing Classroom Management. In D.L. Duke (Ed.), Classroom Management, Seventy-eighth Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 1979.

Jones, R. The Prediction of Teaching Efficiency from Objective Measures. Journal of Experimental Education 15 (1946): 85-99.

Joyce, B. & Weil, M. Models of Teaching. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

Kounin, J.S. & Sherman, L.W. School Environments as Behavior Settings. Theory into Practice 18 (1979): 145-149.

Kozol, J. Death at an Early Age. Boston: Houghton Mifflin, 1967.

Kratz, H.E. Characteristics of the Best Teachers as Recognized by Children. Pedagogical Summary 3 (1896): 413-418.

Levin, T. & Lewy, A. Classroom Environment Study Research Proposal. IEA Secretariat (University of Stockholm), 1978.

Livingstone, D.W.; Hart, D.J. & Davie, L.E. Public Attitudes Toward Education in Ontario. Toronto: OISE Press, 1985.

Mandeville, G.K. Reanalyzing Teaching Research Data: Promises and Pitfalls. Evaluation in Education: An International Review Series 8 (1984): 153-166.

McCall, W. & Krause, G. Measurement of Teacher Merit. Journal of Educational Research 53 (1959): 73-5.

McDonald, F.J. & Elias, P. The Effects of Teaching Performance on Pupil Learning, Beginning Teacher Evaluation Study: Phase II, Final Report, Vol. 1. Princeton, N.J.: Educational Testing Service, 1976.

Medley, D.M. Teacher Effectiveness. In H.E. Mitzel (Ed.), Encyclopedia of Educational Research (5th Ed.). New York: The Free Press, 1982.

Medley, D.M. The Effectiveness of Teachers. In P. Peterson & H. Walbert (Eds.), Research On Teaching. Berkeley, Calif.: McCutchan, 1979.

Medley, D.M. Teacher Competence and Teacher Effectiveness. Washington, D.C.: American Association of Colleges for Teacher Education, 1977.

Mitzel, H.E. Teacher Effectiveness. In C.W. Harris (Ed.), Encyclopedia of Educational Research (3rd Ed.). New York: Macmillan, 1960.

Peterson, P. & Clark, C. Teachers' Reports of Their Cognitive Processes During Teaching. American Educational Research Journal 15 (1978): 555-565.

Purves, A.C. The IEA Empire: A Memoire. Comparative Education Review (in press).

Purves, A.C. Literature Education in Ten Countries: An Empirical Study, International Studies in Evaluation, Vol. 2. New York: John Wiley, 1973.

Romberg, T.A. & Price, G.G. Curriculum Implementation and Staff Development as Cultural Change. In G.A. Griffin (Ed.), Staff Development, Eighty-second Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1983.

Rosenshine, B. Teaching Functions in Instructional Programs. The Elementary School Journal 83 (1983): 335-351.

Rosenshine, B. Content, Time, and Direct Instruction. In P. Peterson & H. Walberg (Eds.), Research on Teaching. Berkeley, Calif.: McCutchan, 1979.

Rosenshine, B. Classroom Instruction. In N.L. Gage (Ed.), The Psychology of Teaching Methods, Seventy-fifth Yearbook of the National Society for the Study of Education, Part I. Chicago: University of Chicago Press, 1976.

Rosenshine, B. Teaching Behaviors and Student Achievement. London: National Foundation for Educational Research, 1971.

Rosenshine, B. The Stability of Teacher Effects upon Student Achievement. Review of Educational Research40 (1970): 647-662.

Rosenshine, B. & Berliner, D.C. Academic Engaged Time. British Journal of Teacher Education 4 (1978) 3-16.

Ryan, D.W. Preactive and Proactive Supervision of Mastery Learning Programs. In D.U. Levine (Ed.), Improving Student Achievement through Mastery Learning Programs. San Francisco: Jossey-Bass, 1985a.

Ryan, D.W. Research on Teaching and Teacher Education: Changing Views on Desirable Research Design and Method. In I.E. Housego & P.P. Grimmett (Eds.), Teaching and Teacher Education: Generating and Utilizing Valid Knowledge for Professional Socialization. Vancouver: University of British Columbia, Faculty of Education, 1985b.

Ryan. D.W. Improving the Content and Process of Pre- and In-Service Teacher Training. Evaluation in Education: An International Review Series 8 (1984): 113-131.

Ryan, D.W. Classroom Observation Strategies. Paper presented at the Conference on Refining Education, Laurentian University, 1981.

Ryan, D.W. & Anderson, L.W. (Eds.), Rethinking Research on Teaching: Lessons Learned from an International Study, special issue of Evaluation in Education: An International Review Series 8 (1984).

Ryan, D.W. & Anderson, L.W. The IEA Classroom Environment Study: Promises and Pitfalls. Evaluation in Education: An International Review Series 8 (1984): 87-93.

Ryan, D.W.; Hildyard, A.; Jantzi, D.; Pike, R.; Hanna, G. & Quinlan, F. The Ontario IEA Classroom Environment Study. Final Report to the Ontario Ministry of Education, 1984.

Ryan, D.W. & Schmidt, M. Mastery Learning: Theory, Research and Implementation. Toronto: Ontario Ministry of Education, 1979.

Silberman, C. Crisis in the Classroom: The Remaking of American Education. New York: Random House, 1970.

Slavin, R.E. & Karweit, N. Cognitive and Affective Outcomes of an Intensive Study Team Learning Experience. Journal of Experimental Education 50 (1981): 29-35.

Smith, L. & Geoffrey, W. The Complexities of an Urban Classroom. New York: Holt, Rinehart & Winston, 1968.

Soar, R.S. Final Report, Follow Through Classroom Process Measurement and Pupil Growth (1970-1971). Gainesville: Institute for the Development of Human Resources, University of Florida, 1973.

Soar, R.S. Teacher-Pupil Interaction. In J.R. Squire (Ed.), A New Look at Progressive Education. Washington, D.D.: Association for Supervision and Curriculum Development, 1972.

Stallings, J. Implications from the Research on Teaching for Teacher Preparation. In I.E. Housego & P.P. Grimmett (Eds.), Teaching and Teacher Education: Generating and Utilizing Valid Knowledge for Professional Socialization. Vancouver: University of British Columbia, Faculty of Education, 1985.

Stallings, J.; Needels, M. & Stayrook, N. How to Change the Process of Teaching Basic Reading Skills in Secondary Schools: Phase II and Phase III; Final Report. Menlo Park, Calif.: SRI International, 1979.

Stallings, J. & Kaskowitz, D. Follow Through Classroom Observation Evaluation, 1972-1973 (SRI Project URU-7370). Menlo Park, Calif.: SRI International, 1974.

Stephens, J. The Process of Schooling. New York: Holt, Rinehart & Winston, 1967.

Stones, E. Practice Teaching. In T. Husen & T.N. Postlethwaite (Eds.), International Encyclopedia of Education. Oxford: Pergamon Press, 1984.

Thorndike, R.L. Reading Comprehension Education in Fifteen Countries, International Studies in Evaluation, Vol. 3. New York: John Wiley, 1973.

Veldman, D. & Brophy, J. Measuring Teacher Effects on Pupil Achievement. Journal of Educational Psychology 66 (1974): 319-324.

Walberg, H.J. A Nation at Risk: Retrospect and Prospect. In T.M. Tomlinson & H. Walberg (Eds.), Academic Work and Educational Excellence. Berkeley, Calif.: McCutchan, 1986.

APPENDIX A

CONSTRUCTS, VARIABLES, AND INSTRUMENTS

Construct/Variables	Instruments
COMMUNITY CHARACTERISTICS	
School Location	ITQ
SCHOOL CHARACTERISTICS	
Type of School (TYPESC)	ITQ
Grade Level Organization	ITQ
Size of School	ITQ
Time Allocated to Subject (Per Year)	ITQ
TEACHER CHARACTERISTICS	
Sex (SEXT)	ITQ
Age (AGET)	ITQ
Years of Experience (YRS)	ITQ
Certification	ITQ
Subject Area Training	ITQ
Subject Contact Hours (Per Week)	ITQ
Number of Additional Subjects Taught	ITQ
Total Contact Hours (Per Week)	ITQ
HOME CHARACTERISTICS	
Father's Education	ISQ
Mother's Education	ISQ
Father's Occupation	ISQ
Language Used in Home	ISQ
STUDENT CHARACTERISTICS	
Sex (SEXS)	ISQ
Age (AGES)	ISQ
Pretest (PRETEST)	PRETEST
Aspirations (ASPIRATION)	ISQ

Attitude Toward School (ATTSCH)	ISQ
Attitude Toward Subject (PREATTSUBJ)	ISQ
CLASSROOM CONTEXT	
Typical Use of Grouping	ITQ
Observed Use of Grouping	FMI
Use of Types of Materials	ITQ
Student Access to Books Outside School	ITQ
Lesson Emphasis	POI
Lesson Objective	POI
Class Size	CII
Number of Adults Per Classroom	CII
Student Attendance	CII
TEACHER PERCEPTIONS	
Teacher Autonomy	ITQ
Relative Class Ability Level	ITQ
Percent of Students Needing Remediation	ITQ
QUANTITY OF INSTRUCTION	
Amount of Homework	ITQ
Time for Homework	ITQ
Opportunity to Learn	OTL
Allocated Time	CII
Instructional Time	FMI
ORIENTING STUDENTS	
Specifying Objectives	ITQ
Reviewing Content	SNAP
TEACHING STUDENTS	
Teacher Interacting	SNAP
Teacher Monitoring	SNAP
Monitoring Individual Seatwork	ITQ
Monitoring Groupwork	ITQ
Lecture	SNAP

Discourse/Discussion	SNAP
Oral Practice/Drill	SNAP
Reading at Seats	SNAP
Written Seatwork	SNAP
Laboratory Work	SNAP
Direction of Verbal Interactions	FMI
Explanation	FMI
Explanation Using Materials	FMI
Demonstration	FMI
Use of Examples	FMI
Structuring Cues	FMI
Directives	FMI
Student Questions	FMI
Student Contributions	FMI
Social Interactions	FMI
Silence (Absence of Verbal Interactions)	FMI
Frequency of Extra Help	FTQ
Number Receiving Extra Help	FTQ
Timing of Extra Help	FTQ
Format of Extra Help	FTQ
Type of Extra Help	FTQ

ASSESSING & EVALUATING STUDENTS

Questioning Student	FMI
Questioning Group	FMI
Asks Recall Questions	FMI
Asks Higher-Order Questions	FMI
Asks Opinion Questions	FMI
Asks Effectiveness Questions	FMI
Amount of Questioning	FTQ
Waits for Student(s) Response	FMI
Student Short Response	FMI
Student Recitation	FMI
Student Extended Response	FMI
Student Doesn't Answer	FMI
Positive Acknowledgement	FMI

Says Answer Wrong	FMI
Repeats Answer	FMI
Redirects Question	FMI
Gives Answer	FMI
Probes	FMI
Probes to Student	FMI
Probes to Group	FMI
Tests for Grading Purposes	FTQ
Diagnostic Testing	FTQ
MANAGING STUDENTS	
Management & Routine Administration	FTQ, SNAP
Procedural Interactions	FMI
Transitional Interactions	FMI
Disciplinary Actions	FTQ, FMI
Uninvolved Teacher	SNAP
STUDENT PERCEPTIONS	
Task-Orientation (TASK)	FSQ
Structure of Lesson (STRUCTURE)	FSQ
Teacher Feedback (FEEDBACK)	FSQ
STUDENT PARTICIPATION	
Percent of Students Engaged in Learning (ENGP)	IAR, SNAP
STUDENT OUTCOMES	
Student Achievement (POSTTEST)	POSTTEST
Attitude Toward Subject (POSTATTSUBJ)	FSQ